Identification of a Marker Indicative of Dairy Faecal Contamination in the Environment

Stephanie Adele Harvey

A Thesis presented in partial fulfilment of the

requirements for the degree of

Master of Science

Massey University, Palmerston North

New Zealand

2001

ABSTRACT

Faecal contamination of aqueous environments remains a significant environmental problem. Faecal pollution causes degradation of both chemical and microbial water quality, exposing the public to a variety of pathogenic organisms. Contamination may originate from direct and indirect faecal sources. Direct sources of faecal contamination include piggery and dairy pond effluents, sewage, birds, bathers, or grazing animals with direct access to waterways. Indirect sources may comprise agricultural run-offs, or leaking septic tanks or sewage distribution systems.

Discrimination of the origin of faecal contamination would enable monitoring agencies to predict associated disease risks more accurately, as well as implement strategies to mitigate the contaminating source. *Escherichia coli,* faecal coliforms and enterococci are extensively used as indicators of faecal pollution in water. These organisms are, however, widely distributed in the intestines of warm-blooded animals and hence enumeration does not define faecal origin.

It has been suggested that bacterial strains may adapt to their environmental niche and consequently, host association of strains may be apparent. The development of highly discriminatory molecular fingerprinting techniques provides an opportunity to distinguish closely related strains. These techniques may be suitable for locating host associative factors and hence defining the source of faecal contamination.

The scope the research described in this thesis was to develop a method capable of discriminating sources of faecal contamination in the environment. Agricultural faecal sources were primarily targeted as they represent a significant source of faecal contamination in New Zealand waterways. Two molecular fingerprinting techniques — Randomly Amplified Polymorphic DNA and Amplified Fragment Length

Polymorphism — were investigated for their ability to detect genotypic markers in *E. coli.*

A polymorphic fragment (714 bp) indicative of dairy cattle faecal isolates was identified by AFLP analysis. To facilitate rapid screening of isolates, PCR primers were designed to amplify a segment (462 bp) of the polymorphic fragment. The marker was specific for dairy cattle faecal isolates and was present in approximately half of the strains. Field study results demonstrated that the marker provided a feasible approach for monitoring "special interest" samples such as monitoring significant pollution incidences, supporting prosecution cases or identifying an illusive source of persistent water quality degradation at a particular site. In these instances the diagnostic marker may assist by verifying or eliminating suspected contaminating sources.

AFLP analysis was used to locate a marker diagnostic of faecal origin, and indicated that further markers could be identified. Although RAPD-PCR analysis did not locate a diagnostic marker, the technique appeared to corroborate the AFLP results. The inability to obtain a highly specific marker using RAPD analysis may have been a function of the limited primer combinations that were screened. The AFLP technique could be used to construct a library of markers, enabling differentiation of a wide range of contaminating sources.

ACKNOWLEDGEMENTS

I would like to gratefully thank Dr George Ionas for his supervision of this research, especially for his enthusiasm and direction over the long haul of undertaking a part-time Masters thesis.

Thanks too, to the Protozoa lab, for sharing their benches and to Trish McLenachan for her enthusiasm and assistance with the AFLP technique.

My sincere appreciation must go to Environment Waikato and Dr Tony Petch, for funding the research and thereby enabling me to undertake a project that was both interesting and commercially relevant. I hope that you find the results as interesting as I found the research.

Thanks too, to my colleagues from NZDRI (Dr's Tim Coolbear, Julian Reid, Jean Garman, Paul McJarrow and many others) and my close friends (Cath, Darryl, Sophia and Vanessa) for their support and encouragement over the past few years and especially their patience over the last few months. Thanks also to my families (the Thomson and Harvey clans), for their strong belief in education, their support and understanding.

And without a doubt, my heart felt appreciation must go to my husband and closest friend, Shane, for his perseverance, and endless encouragement and support - Kia kaha my friend it's your turn next (Thanks Tuti for keeping him sane!).

Lastly and most importantly I thank God for ultimately providing me with the opportunity to undertake this research, for providing fantastic people to support and encourage me, and for helping me to keep the project in perspective — to remember that there is much more to life than the deadline which is looming ahead!

TABLE OF CONTENTS

Abstract i

Acknowledgements iii

Table of Contents iv

List of Figures xii

List of Tables xiv

Abbreviations xv

1 INTRODUCTION 1

1.1 SCOPE 1

1.2 WATER POLLUTION AND HEALTH 1

1.3 DETECTING FAECAL CONTAMINATION IN WATER 3

1.4 DIFFERENTIATING FAECAL CONTAMINATION 4

1.5 CHEMICAL CONSTITUENTS OF FAECAL CONTAMINANTS 5

1.5.1 Faecal Sterols 5

1.5.2 Fluorescent Whitening Agents and Sodium Tripolyphosphate 6

1.6 MICROBIAL CONSTITUENTS OF FAECAL CONTAMINANTS 6

1.6.1 Species Differences 6

1.6.1.1 Rhodococcus 7

1.6.1.2 Bifidobacteria. 7

1.6.1.3 Faecal Streptococci Species Profiles 8

1.6.1.4 Faecal Coliform: Faecal Streptococci Ratio 9

1.6.1.5 Bacteriodes Species 10

1.6.1.6 Bacteriodes fragilis Bacteriophage 11

1.6.1.7 F^+ RNA Coliphage Serotypes 12

1.6.1.8 Summary of Differentiation Based on Species Composition 13

1.6.2 Phenotypic Differentiation 14

1.6.2.1 Antibiotic Resistance 14

1.6.2.2 Biochemical Patterns 17

1.6.2.3 Summary of Phenotypic Differentiation 18

1.6.3 Genotypic Differentiation 18

1.6.3.1 *REP & ERIC primers* 18
1.6.3.2 *RAPD-PCR: Location of a Marker Diagnostic of Sewage* 19
1.6.3.3 *Summary of Genotypic Differentiation* 21
1.7 *ESCHERICHIA COLI* 21
1.7.1 Introduction 21
1.7.2 *E. coli* as an Indicator 22
1.7.3 Host Specificity in *E. coli* 23
1.8 LOCATING DIFFERENTIAL MARKERS 25
1.8.1 Genotyping Methods 25
1.8.1.1 *Randomly Amplified Polymorphic DNA* 26
1.8.1.2 *Amplified Fragment Length Polymorphism* 26
1.9 RESEARCH AIMS 28

2 SURVEY FOR FAECAL SPECIFIC MARKERS – RAPD 29

2.1 INTRODUCTION 29
2.1.1 RAPD-PCR Technique 29
2.1.2 Research Strategy 30
2.2 MATERIALS 31
2.2.1 Chemicals 31
2.2.2 Reagents 31
2.2.2.1 *Buffer-saturated Phenol, pH 7.9* 31
2.2.2.2 *10% CTAB* 31
2.2.3 Enzymes, DNA Ladders and Primer Synthesis 31
2.2.4 Media 31
2.3 METHODS 32
2.3.1 Media Preparation 32
2.3.1.1 *Sterilisation* 32
2.3.1.2 *Diluent* 32
2.3.1.3 *Luria Broth-Millers; mFC Agar; Nutrient Agar with MUG* 32
2.3.1.4 *Luria Agar-Millers* 32
2.3.2 Sample Collection 33
2.3.3 Isolation of Faecal *Escherichia coli* 33
2.3.4 Isolates Used to Screen for Markers 34
2.3.5 Genomic DNA Miniprep 34

2.3.6 Agarose Gel Electrophoresis 35
2.3.7 Verification of *E. coli* 35
2.3.8 RAPD-PCR Amplification 36
2.4 RESULTS 37
2.4.1 Verification of *E. coli* Isolates 37
2.4.2 Genotypic Fingerprints Generated by RAPD-PCR 37
2.4.3 Putative Genus-Specific Markers 38
2.4.4 Diagnosis of Faecal Origin 38
2.5 DISCUSSION 40
2.5.1 Application of RAPD-PCR Analysis for Locating Diagnostic Markers 40
2.5.1.1 Putative Swine Marker 40
2.5.1.2 Differentiating Human and Animal Isolates 41
2.5.2 Conclusions 42
2.5.3 Future Research 43

3 IDENTIFICATION OF FAECAL SPECIFIC MARKERS – AFLP 45

3.1 INTRODUCTION 45
3.1.1 Amplified Fragment Length Polymorphism Analysis 45
3.1.2 Research Strategy 49
3.2 MATERIALS 49
3.2.1 Chemicals 49
3.2.2 Reagents 49
3.2.2.1 5 x Restriction Enzyme Buffer 49
3.2.2.2 1 x TBE Buffer 50
3.2.2.3 Silver Staining Solution 50
3.2.3 Enzymes, DNA Ladders, Primer Synthesis, Restriction Endonucleases 50
3.3 METHODS 50
3.3.1 Sterilisation 51
3.3.2 Sample Collection and Isolation of *E. coli* Strains 51
3.3.3 AFLP Analysis 51
3.3.3.1 Restriction Enzyme Digestion 51
3.3.3.2 Ligation of Oligonucleotide Adpaters 51
3.3.3.3 Pre-selective PCR Amplification 52
3.3.3.4 Selective PCR Amplification 52

3.3.3.5 Resolution of AFLP Products 54
3.3.4 AFLP Reproducibility 55
3.4 RESULTS 56
3.4.1 Reproducibility of AFLP Genomic Fingerprints 56
3.4.2 Resolution of Genomic Fingerprints 56
3.4.3 Locating Markers by AFLP 57
3.4.3.1 Putative Genus Specific Marker 57
3.4.3.2 Differentiation of Strains 58
3.4.3.3 Putative Marker of Dairy Faecal Strains 59
3.5 DISCUSSION 60
3.5.1 Resolution of AFLP Products 60
3.5.2 Application of AFLP to Locating Specific Faecal Markers 61
3.5.2.1 Putative E. coli *Diagnosis* 61
3.5.2.2 Differentiation of Strains 62
3.5.2.3 Diagnosing Faecal Origin 63
3.5.3 Host specific Association of Strains 64
3.5.4 Comparison of Techniques – AFLP versus RAPD Analysis 64
3.5.5 Conclusions 65
3.5.6 Future Research 66
3.5.6.1 Dairy Cattle Faecal Marker 66
3.5.6.2 Alternative Markers 67

4 PCR PRIMER DESIGN FOR DETECTION OF THE DAIRY FAECAL MARKER **68**

4.1 INTRODUCTION 68
4.1.1 Research Strategy 68
4.2 MATERIALS 69
4.2.1 Chemicals 69
4.2.2 Reagents 69
4.2.2.1 Buffered phenol, pH 7.9 69
4.2.2.2 STET Buffer 69
4.2.3 Enzymes, DNA Ladders and Primer Synthesis 69
4.2.4 Media 69
4.3 METHODS 70

4.3.1 Media Preparation 70
4.3.1.1 Sterilisation 70
4.3.1.2 LB Broth 70
4.3.1.3 LB Plates (containing amp/IPTG/X-gal) 70
4.3.2 Sequence Analysis of the Putative Dairy Faecal Markers 71
4.3.2.1 Location of the Putative Swine Faecal Marker 71
4.3.2.2 Location of the Putative Dairy Faecal Marker 71
4.3.2.3 Band Elution from Agarose Gels 71
4.3.2.4 Band Elution from Polyacrylamide Gels 71
4.3.2.5 Transformation of Putative Marker 72
4.3.2.6 Plasmid DNA Extractions 72
4.3.2.7 Agarose Gel Electrophoresis 73
4.3.2.8 DNA Purification 73
4.3.2.9 Sequencing Analysis of the Insert 73
4.3.3 Target-Specific PCR for Putative Dairy Faecal Marker 74
4.3.3.1 Primer Design 74
4.3.3.2 DNA Lysate Preparation - Rapid Boil Method 74
4.3.3.3 Target-specific PCR for the Putative Dairy Faecal Marker 75
4.3.3.4 Sequence Verification of Amplified PCR Product 75
4.3.4 Restriction Enzyme Digestion of Amplified Dairy Faecal Marker 76
4.3.5 Marker Specificity of the Putative Dairy Faecal Marker 76
4.4 RESULTS 77
4.4.1 Sequence Analysis of the Putative Faecal Markers 77
4.4.1.1 Fragment Transformation 77
4.4.1.2 Sequence Analysis of the Putative Swine Marker 77
4.4.1.3 Identity of the Putative Dairy Faecal Marker 77
4.4.2 Target-specific PCR for Putative Dairy Cattle Faecal Marker 78
4.4.3 Restriction Enzyme Digestion 79
4.4.4 Specificity of the Putative Dairy Faecal Marker 80
4.4.5 Marker Stability Over Time 81
4.5 DISCUSSION 82
4.5.1 Location of the Polymorphism in Dairy Faecal Strains 82
4.5.2 Target-specific PCR 84
4.5.3 Specificity of the Dairy Faecal Marker 85
4.5.3.1 Differentiation of Dairy and Beef Cattle 86
4.5.4 Application to Environmental Monitoring 87

4.5.5 Conclusions 88
4.5.6 Future Research 88

5 FIELD EVALUATION - DAIRY FAECAL MARKER 90

5.1 INTRODUCTION 90
5.2 MATERIALS 90
5.2.1 Chemicals 90
5.2.2 Enzymes, DNA Ladders and Primer Synthesis 90
5.2.3 Media 91
5.3 METHODS 91
5.3.1 Media Preparation 91
5.3.1.1 Sterilisation 91
5.3.1.2 Diluent 91
5.3.1.3 Preparation of Media 91
5.3.2 Reproducibility of Marker Detection 91
5.3.2.1 Replicate Sample Analysis 91
5.3.2.2 Replicate PCR Detection of a Set of Isolates 92
5.3.2.3 Isolation of Faecal Escherichia coli 92
5.3.2.4 PCR Amplification-Optimising High Throughput Screening 92
5.3.2.5 Agarose Gel Electrophoresis 93
5.3.3 Field Study 93
5.4 RESULTS 94
5.4.1 PCR Amplification using Whole Cells versus Lysate DNA 94
5.4.2 Reproducibility of Marker Detection 94
5.4.2.1 Replicate Sample Analysis 94
5.4.2.2. Replicate PCR Detection from a Set of Isolates 95
5.4.3 Field Study 95
5.5 DISCUSSION 97
5.5.1 Whole Cell Suspensions versus DNA Lysates 97
5.5.2 Replicate Analysis of Sample 98
5.5.3 Field Study 99
5.5.3.1 Feasibility of the Detection System 99
5.5.3.2 Diagnosis of Faecal Source 99
5.5.3.3 Resolving the Inconclusive Results 102

5.5.4 Conclusions 103
5.5.5 Future Research 103

6 GENERAL CONCLUSIONS 104

6.1 MOLECULAR PROFILING FOR MARKER LOCATION 104
6.1.1 Differentiating Sources of Faecal Contamination 105
6.1.1.1 Dairy Cattle Faecal Marker 105
6.1.1.2 Animal Markers 106
6.1.1.3 Other Markers 106
6.1.1.4 Differentiation Based on Genotypic Fingerprints 106
6.2 FUTURE RESEARCH 107
6.2.1 Recommendations for Monitoring Prevalence of the Dairy Cattle Faecal Marker 107
6.2.2 Recommendations for Further Markers 109
6.2.3 Discrimination of Faecal Source using Genotypic Profiles 109

REFERENCES 109

APPENDIX 132

A REAGENTS 132
A.1 1 x E Buffer 132
A.2 1 x TE Buffer 132
A.3 Loading Dye for Agarose Gels 132
A.4 Formamide Loading Dye for Polyacrylamide Gels 132
A.5 AFLP Adapter Preparation 132

B PCR OPTIMISATION OF THE DAIRY FAECAL MARKER 133
B.1 $MgCl_2$ Optimisation 133
B.1.1 Method 133
B.1.2 Results & Conclusions 133
B.2 Primer Annealing Temperature 135

B.2.1 *Method* 135
B.2.2 *Results & Conclusions* 135
B.3 Number of PCR Cycles 137
B.3.1 *Method* 137
B.3.2 *Results & Conclusions* 137
B.4 *Taq* Polymerase Brand 138
B.4.1 *Method* 138
B.4.2 *Results & Conclusions* 138
B.5 Summary- Optimised TS-PCR Conditions For Dairy Faecal Marker 139

C SEQUENCE DATA 140
C.1 pGEM®-T Easy Vector Sequence 140
C.2 Alignment of Dairy Polymorphism with *E. coli* K12 141
C.3 PCR Primers Designed for Dairy Marker 142
C.4 *Pst* I Digestion of the 462 bp Dairy Faecal Marker 143

LIST OF FIGURES

Chapter Two – RAPD-PCR Analysis

Figure 2.1 *E.coli* verification of isolates 37

Figure 2.2 Genotypic fingerprints of faecal *E.coli* isolates using primer GC70 38

Figure 2.3 Genotypic fingerprints of faecal *E.coli* isolates using primer GC60 39

Chapter Three – AFLP-PCR Analysis

Figure 3.1 AFLP Schematic 46

Figure 3.2 Genotypic fingerprints of *E.coli* isolates using AFLP primers E-GA/M-CA 57

Figure 3.3 Genotypic fingerprints of *E.coli* isolates using AFLP primers E-AT/M-AG 58

Figure 3.4 Genotypic fingerprints of *E.coli* isolates using AFLP primers E-AGC/M-CT 59

Figure 3.5 Genotypic fingerprints of *E.coli* isolates using AFLP primers E-ATT/M-CTA 60

Chapter Four – Method Development-Dairy Faecal Marker

Figure 4.1 Target-specific PCR of the putative dairy faecal marker 79

Figure 4.2 *Pst* I digestion of the dairy faecal marker & non-target product 80

Chapter Five – Evaluation-Dairy Faecal Marker

Figure 5.1 Optimisation of initial denaturation time using whole cell suspensions 94

Figure 5.2 Dairy cattle faecal marker frequency at each of the field sites. 96

Appendix

Figure B.1 Optimisation of $MgCl_2$ concentration for TS-PCR amplification of the dairy marker 134

Figure B.2 Optimisation of primer annealing temperature for TS-PCR amplification of the dairy marker 136

Figure B.3 Optimisation of the number of amplification cycles for TS-PCR amplification of the dairy markers 138

Figure B.4	Comparison of *Taq* polymerase brand for TS-PCR amplification of the dairy marker	139
Figure C.1	pGEM-T Easy Vector Sequence	140
Figure C.2	Alignment of dairy polymorphism and *E. coli* K12	141
Figure C.3	PCR primers design for dairy marker	142
Figure C.4	*Pst*I digestion of the 462 bp dairy faecal marker	143

LIST OF TABLES

Chapter Two – RAPD-PCR Analysis

Table 2.1 Oligonucleotide primers 32

Table 2.2 Geographical distribution of faecal samples 33

Chapter Three – AFLP-PCR Analysis

Table 3.1 AFLP adapters sequences 50

Table 3.2 Pre-selective primers for AFLP amplification 52

Table 3.3 AFLP-PCR amplification conditions 53

Table 3.4 Selective primer combinations for AFLP amplification 53

Chapter Four – Method Development-Dairy Faecal Marker

Table 4.1 Sequencing and target-specific PCR primers 70

Table 4.2 Specificity of putative dairy faecal marker – dairy faecal isolates 81

Table 4.3 Specificity of putative dairy faecal marker – non-dairy faecal isolates 82

Chapter Five – Evaluation-Dairy Faecal Marker

Table 5.1 Oligonucleotide primers for the dairy cattle faecal marker 91

Table 5.2 Replicate analysis of marker positive isolates 95

Table 5.3 Enumeration of faecal coliforms, *E. coli* and marker positive isolates at each of the field sites 97

ABBREVIATIONS

AFLP	Amplified Fragment Length Polymorphism
amp	ampicillin
bp	base pairs
CTAB	hexadecyltrimethylammonium bromide
E. coli	*Escherichia coli*
Eco RI	Restriction endonuclease with 6 base recognition site
ERIC-PCR	Enterobacterial Repetitive Intergenic Consensus Sequence amplified by the Polymerase Chain Reaction
F^+ RNA phage	RNA coliphage which attaches to the F-pili of bacterial strains
IPTG	isopropyl-β-D-thiogalactosidase
LB	Luria broth for transformant strains
mFC	Agar for the selective cultivation of faecal coliforms
Mse I	Restriction endonuclease with 4 base recognition site
MUG	4-methyl-umbelliferyl-β-D-glucuronide
PCR	Polymerase Chain Reaction
pGEM®-T Easy Vector	Vector for cloning PCR products
Pst I	Restriction endonuclease with 6 base recognition site
RAPD	Randomly Amplified Fragment Length Polymorphism
REP-PCR	Repetitive Extragenic Palindromic amplified by the Polymerase Chain Reaction
STET	Saline, Tris-HCl, EDTA, Triton X-100 buffer
TE	Tris-EDTA buffer
TS-PCR	Target-specific Polymerase Chain Reaction
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

1 Introduction

1.1 SCOPE

The aim of the research described in this thesis was to locate a marker(s) indicative of the origin of faecal contamination in waterways. These molecular markers would enable monitoring agencies to more accurately trace the cause of the pollution and hence increase the opportunity to remediate the problem. Two molecular methodologies were investigated, Randomly Amplified Polymorphic DNA and Amplified Fragment Length Polymorphism, to detect genetic differences in *Escherichia coli* from humans and animals.

1.2 WATER POLLUTION AND HEALTH

The presence of pathogenic bacteria, viruses and parasitic protozoa are the most common and widespread health problem inherent to drinking water. According to McFeters & Singh (1991), waterborne enteric disease remains "a major source of human mortality and morbidity", with over half of the world's population contracting some form of waterborne illness resulting from polluted water. Although it is acknowledged that the vast majority of incidences occur in developing countries, the last decade has shown an increasing trend in industrialised nations (McFeters & Singh, 1991). Although the former may experience frequent isolated disease outbreaks, the latter in contrast, may intermittently experience simultaneous infection of large populations (Bonde, 1977; Davis, 1971).

Pathogens in the environment often originate from the faecal matter of humans or animals carrying the disease-causing organism. Over one hundred different viral types have been identified in domestic sewage (Tartera & Jofre, 1987). The survival, proliferation and virulence of the viruses vary from organism to organism. Pathogens commonly infecting humans are those that either survive for long periods following excretion, or are highly infective, such that ingestion of a few cells will

result in disease (Ministry of Health, 1995). Pathogens capable of surviving in aqueous environments consequently have the potential to be rapidly and widely dispersed (Ministry of Health, 1995). Consequently, human infection *via* water transmission is often more significant than infection resulting from person to person (or animal) contact (Bonde, 1977).

Pathogenic organisms transmitted by water proliferate predominantly in the intestinal tract, where they may multiply to high densities, cause health related problems followed by excretion in the faeces and so exacerbate the problem (Ministry of Health, 1995). The contamination of aqueous environments by either animal or human faecal matter may cause significant community problems, as faecal matter may enter water supplies or areas of recreational use, and subsequently expose users to the risk of waterborne disease. Adequate treatment of drinking water supplies ensures the efficient removal of the majority of pathogens, providing "safe" potable water. However, if breakdowns occur during treatment processes or, alternatively, if supplies are untreated prior to consumption, there is potential that pathogens may cause infection.

Transmission is also possible during bathing in contaminated waters, as shown by correlation between the incidence of gastro-enteritis and swimmers using contaminated bathing water (Bitton, 1994; Cabelli, 1989; Sinton *et al.,* 1993b). Consequently, health hazards associated with contaminated waterways are of considerable importance to society. Extensive monitoring of drinking and recreational water are undertaken by various agencies throughout the world to assess the risk of health hazards, enabling corrective actions which minimise detrimental impacts on users. Within New Zealand itself, 19% of people appraised in the 1998 Microbiological Quality of Drinking Water Survey, were supplied water that failed the faecal coliform criteria (Ball, 1999). Although a number of reasons were responsible, the figure illustrates that microbiological quality of water remains a significant issue in environmental and health management.

1.3 DETECTING FAECAL CONTAMINATION IN WATER

Benchmark studies in the 1850's, established a correlation between cholera and faecal contamination of drinking water (see: Edberg *et. al.,* 2000). Subsequent research has attempted to identify markers indicative of faecal contamination, *via* either direct detection of pathogenic organisms or indirect detection methods employing indicator organisms (Ministry of Health, 2000).

Direct detection of pathogens is not routinely utilised for monitoring purposes as detection is attenuated due to the extensive and tedious methodologies of conventional protocols (APHA, 1995). Recent developments utilising PCR amplification and DNA hybridisation have improved detection methods - *e.g.* detection of Hepatitis A (Lees *et al.,* 1995), enteroviruses (Abbaszadegan *et al.,* 1993) and *Giardia* cysts (Mahbubani *et al.,* 1991). However the usefulness and feasibility of direct detection protocols continue to be limited by the variety of pathogens potentially present, their intermittent temporal appearance and insufficient sensitivity of the detection techniques (Bonde, 1977; Ministry for the Environment, 1997; Ministry of Health, 2000; Sinton *et al.,* 1993a; West, 1991).

For these reasons monitoring agencies have predominantly employed surrogate indicators to monitor microbiological water quality. Such indicator organisms are typically non-pathogenic but are instead synonymous with the presence of pathogens, as their originating source is identical (Sinton *et al.*, 1993a). No single indicator group is capable of providing sufficient disease risk indication in all situations, but rather a combination of indication systems is recommended (Grabow, 1986; Ministry for the Environment, 1997; Sinton *et al.*, 1993a). Current methods for the assessment of water quality utilise enumeration of enterococci as the primary indicator of faecal contamination in marine recreational waters and *E. coli* in fresh water (APHA, 1995). Both the enterococci group and *E. coli* are, however, present in a variety of warm-blooded animals and are therefore unsuitable for indicating the actual origin of any faecal contamination (Fiksdal *et al.,* 1985; APHA, 1995; Sinton & Donnison, 1994).

The requirements for an effective water quality indicator are: (1) that the organism must be universally present in high numbers in the faecal matter of humans and animals, and it must be readily detected by simple and inexpensive methods. (2) The indicator organism should not grow or multiply in the external environment, but its survival should ideally represent that of the pathogens of which it is a surrogate indicator (Edberg *et al.*, 2000).

1.4 DIFFERENTIATING FAECAL CONTAMINATION

Animals and humans may be infected with different pathogens and transmission across the two groups is known to occur (Bielaszewska *et. al.,* 2000; Brown 1999; Johnson *et al.,* 2000; Hooper, 1999; Linton, 1977; Swayne, 1999). Until recently, it was thought that some of the most significant human pathogens originated from other infected humans. As such, many waterbourne diseases were almost exclusively attributed to water supplies polluted by human sewage (Jagals *et al.,* 1995). When faecal contamination consists of animal or stormwater runoffs rather than from human origin, the prediction of disease risk using indicator bacteria was thought to be less accurate (Ministry for the Environment, 1997). This affect may, however, be localised, as stormwater runoffs have been shown to contain significant levels of human faecal contamination, as a result of the age and effectiveness of sewage distribution systems, septic tanks and the presence of illegal sewer connections (Leeming *et al.,* 1998). However, a paradigm shift has emerged over recent years as reported in the Recreational Water Quality Guidelines (Ministry for the Environment, 1999a) – there is no significant difference in the health risks associated with either animal or human faecal contamination.

The advantages of differentiating faecal sources would be two-fold. First, it would enable correlation of certain pathogens with particular faecal sources, thereby enabling more accurate assessment of associated health hazards. Second, knowledge of the pollution type would assist identification of the geographical source of contamination and therefore increase opportunity to remediate pollution incidences. Research attempting to provide differentiation has examined both chemical and

microbial constituents of various faecal inputs. The latter concept has targeted niche specificity exhibited at the species level, or the phenotypic or the genotypic level (Dombek *et al.,* 2000; Hagedorn *et al.,* 2000; Sinton, Finlay & Hannah, 1998; Turner, 1996)

1.5 CHEMICAL CONSTITUENTS OF FAECAL CONTAMINANTS

1.5.1 Faecal Sterols

The term "faecal sterols" encompasses a number of compounds including coprostanol (5β-coprostan-3β-ol). This 5β-stanol has been used as a reliable indicator of faecal contamination in the environment even when microbial indicators have been destroyed. Coprostanol is produced exclusively in the intestinal tract of humans and higher mammals through bacterial reduction of cholesterol. Hence the sterol fingerprint represents a combination of diet, metabolism and intestinal flora (Brown & Wade, 1984; Sinton *et al.,* 1998). Not surprisingly, researchers have found that the profile of faecal sterols from different animals varies. The major component of human faecal matter is coprostanol, whereas the sterol profile of herbivores is dominated by 24-ethylcoprostanol, with 5β-stanols in generally equal or greater abundance than 5α-stanols (Leeming *et al.,* 1996; Sinton *et al.,* 1998). The abundance of 5α- and 5β-stanols was low in the sterol profile of dogs and birds. Cats and pigs were the only mammals to exhibit profiles similar to humans.

Leeming *et al.* (1996) used faecal sterol profiles to trace sources of faecal contamination. During wet weather, profiles revealed that indicator bacterial levels were due to human faecal inputs, whereas during dry weather, profiles indicated that bacterial counts did not originate from human or herbivore sources. As such, the protocol appears to provide suitable discrimination of faecal inputs. Detection of faecal sterols requires solvent extraction of samples, followed by alkaline saponification and subsequent sterol identification by gas chromatography linked to mass spectrometry (GC-MS) (Leeming *et al.,* 1996). Although such techniques are not routinely used in New Zealand water quality laboratories, the contracting of

analyses to private laboratories has increased over recent years and hence such methodologies may not be unrealistic.

1.5.2 Fluorescent Whitening Agents and Sodium Tripolyphosphate

Fluorescent whitening agents (FWA) and sodium tripolyphosphate (STP) are constituents of washing powders and are discharged with washing wastewater. Hence they become potential indicators of human sewage (Sinton *et al.,* 1998). Moreover, levels of sodium tripolyphosphate have shown significant correlation ($P < 0.001$) with faecal coliform levels (Close *et al.,* 1989). FWA and STP have been used in New Zealand to indicate septic tank effluent contamination of wells even when other water quality determinands met the chemical criteria for drinking water. The survey indicated that 50% of well samples failed the total coliform criteria for drinking water and that 17% contained either fluorescent whitening agents or sodium tripolyphosphate. STP is however, susceptible to hydrolysis in aqueous environments and sediment adsorption, effectively reducing the concentration in the receiving water, and hence it's effectiveness as an indicator (Sinton *et al.*, 1998).

1.6 MICROBIAL CONSTITUENTS OF FAECAL CONTAMINANTS

1.6.1 Species Differences

Literature on the differentiation of faecal origin is dominated by research examining speciation (Kreader, 1995; Jofre *et al.,* 1989; Mara & Oragui, 1981; Mara & Oragui, 1983; Sinton *et al.,* 1993a; Sinton *et al.,* 1993b). This approach has been largely founded on benchmark studies by Moore & Holdeman (1974), which characterised the species composition of the human faecal microbial population. Subsequent researchers typically targeted species that were present in the human faecal microflora but absent in non-human sources, or *vice versa.* As a consequence *Rhodococcus, Bifidobacterium*, faecal streptococci, *Bacteriodes* and various bacteriophage have been investigated.

1.6.1.1 Rhodococcus

Rhodococcus coprophilus, a nocardioform actinomycete, is a natural inhabitant of the faeces of herbivores. It has been isolated from the faeces of domestic animals (cattle, sheep, pigs, horses), poultry reared in close proximity to herbivores, agricultural run-offs, and associated wastewaters and water courses (Jagals *et al*, 1995; Mara & Oragui, 1981; Oragui & Mara, 1983; Rowbotham & Cross, 1977; Sinton *et al.*, 1998). *R. coprophilus* counts in agricultural streams have been shown to be higher than those of other faecal indicators (Oragui & Mara, 1983). The organism is apparently absent in faecal matter originating from human sources and hence *R. coprophilus* has been suggested as an indicator of non-human faecal contamination.

Although the organism appears specific to non-human faecal matter, several limitations exist for its use as a source-specific indicator. In the first instance, studies have reported that *R. coprophilus* can survive under environmental conditions four to five times longer than *E. coli* (Oragui & Mara, 1983). The longevity of the organism restricts its usefulness to indicating remote farm animal faecal pollution. Combination with *Streptococcus bovis* enumeration could be used to indicate recent faecal contamination, as *S. bovis* survives only a few days. Secondly and most importantly, extended incubation times (16 - 21 days) are required for the cultivation of *R. coprophilus.* Such incubation times significantly delay identification of the contaminating source and its associated health risks, making the use of *R. coprophilus* unsuitable for routine monitoring purposes.

1.6.1.2 Bifidobacteria.

Early reports by Buttiaux and Mossel (1961) (see: Mara & Oragui, 1983) suggested that the genus *Bifidobacterium* would be an ideal faecal indicator. Later studies reported that this genus might be suitable for distinguishing human and animal faecal contamination. Mara & Oragui (1983) showed that sorbitol-fermenting Bifidobacteria (mainly *B. adolescentis* and *B. brevé*) were only isolated from the faecal matter of humans, whereas Bifidobacteria fermenting mannitol were

consistently isolated from both animals and humans. Subsequent research supported the association of sorbitol-fermenting Bifidobacteria with isolates of human origin (Jagals & Grabow, 1996; Jagals *et al*, 1995). Although Resnick and Levin (1981) reported that Bifidobacteria could be isolated from the faeces of humans and swine, the apparent contradiction may be attributable to variations in the selective media used (Sinton *et al.*, 1998). Recent research employing molecular methods identified host specific markers (human and bovine) in both the genus *Bifidobacterium* and the *Bacteriodes-Provetella* group (Bernhard & Field, 2000). Host specific patterns based on 16S rDNA markers suggested that there were differences in the species composition between humans and animals (Bernhard & Field, 2000).

Comparative survival studies between sorbitol-fermenting Bifidobacteria and *E. coli* indicated that the survival of Bifidobacteria was significantly lower - *Bifidobacterium* survival in fresh water was up to 26 times less than that of *E. coli* (Mara & Oragui, 1983). In addition, *Bifidobacterium* species were extremely sensitive to chlorination (Mara & Oragui, 1983). Therefore it was concluded that although Bifidobacteria may enable differentiation of the origin of faecal contamination, their usefulness as indicators of potential pathogenic presence is limited to recent contamination in non-chlorinated waters (Jagals *et al.*, 1995; Mara & Oragui, 1983). Moreover, Bifidobacteria are strict anaerobes and hence require specialised equipment and expertise for anaerobic cultivation of strains. Furthermore, there has been little field evaluation of the genus and hence there are no epidemiological data or water quality standards that correlate sorbitol-fermenting Bifidobacteria with disease-risk.

1.6.1.3 Faecal Streptococci Species Profiles.

Another differential approach has examined the species composition of faecal streptococci in a variety of faecal extracts and effluents. The faecal streptococci originate from the faeces of warm-blooded animals and are generally considered to comprise of *Streptococcus bovis* and *S. equinus*, as well as *Enterococcus faecium, E. faecalis, E. durans, E. avium* and *E. gallinarum* (Sinton *et al.,* 1993a; Sinton *et al.,* 1998). While the literature appears contradictory, general consensus concerning the

composition of faecal streptococci from humans and animals has been attained. Different proportions of enterococci and non-enterococci species have been shown in the faecal matter of various vertebrates (Sinton & Donnison, 1994; Sinton *et al.,* 1993b). Enterococci species dominate the faecal matter of humans, whereas the streptococococcal population from animals comprises enterococci and significant levels of non-enterococci species. *S. equinus* and *S. bovis* are generally restricted to animals — the latter constituting more than 70% of the streptococcal flora of cattle (Sinton & Donnison, 1994; Sinton *et al.,*1993b; Sinton *et al.,* 1998).

Although limited differentiation of faecal sources may appear feasible, faecal streptococci have been recovered from plants, soil and insects, and therefore are not specific to the faecal matter of vertebrates (Sinton *et al.,* 1993b). In addition, enterococci species survive longer than streptococci, consequently altering the compositional ratios over time. As such, the ratios detected in receiving waters may no longer accurately reflect the faecal origin. Consequently, faecal streptococci do not provide a reliable indicator of faecal origin (Sinton & Donnison, 1994; Sinton *et al.,* 1993b; Sinton *et al.,* 1998).

1.6.1.4 Faecal Coliform: Faecal Streptococci Ratio.

Researchers have also suggested that the ratio of faecal coliforms to faecal streptococci (FC:FS) might be indicative of faecal origin. The ratio relies on the counts of faecal streptococci predominating over those of faecal coliforms in animals, and *vice versa* in humans (Sinton *et al.,* 1993b). In human faeces the FC:FS ratio is > 4, whereas ratios of < 0.7 are obtained from animal faeces (APHA, 1995; Sinton & Donnison, 1994; Sinton *et al.,* 1993b; Sinton *et al.,* 1998).

A number of factors have lead to the abandonment of the FC:FS ratio as an indicator of the origin of faecal contamination in water (APHA, 1995). The two bacterial groups exhibit significantly different survival rates. A continuum of die-off rates in natural waterways may be presented as follows: *S. bovis* - equinus group > faecal coliforms > *E. faecium* - *faecalis* group (Sinton *et al.,* 1993b). Consequently the ratios may alter over time and hence may no longer accurately reflect the origin of

faecal contamination (McFeters *et al.,* 1974). Although other researchers have argued that die-off rates would enhance the ratio (Feachem, 1975), the ratio is also affected by bacterial counts which differ significantly between animals, temperature, sediment presence and sediment particle size (Howell *et al.,* 1995). Furthermore the FC:FS ratios from processing effluents of abattoirs has been shown to be similar to those from sewage (Sinton & Donnison, 1994), and ratios > 4 have been reported in cattle and sheep (Mara & Oragui, 1981). To date the protocol is too unreliable to discriminate human and domestic animal faecal contamination (Howell *et al.,* 1995).

1.6.1.5 *Bacteriodes* Species

Strains belonging to the genus *Bacteriodes* are obligate anaerobes and dominate the faecal microflora of humans (Moore & Holdeman, 1974). With counts 10 - 100 fold higher than that of *E. coli*, the *Bacteriodes* group account for up to 30% of all human faecal isolates (Holdeman *et al.,* 1976; Moore & Holdeman, 1974; Sinton *et al.,* 1998). Researchers have assessed various components of the *Bacteriodes* group as prospective indicators of human faecal contamination. For example, although the presumptive *Bacteriodes fragilis* group (pBFG) dominates the human faecal flora, they are rarely detected, if at all, in the faecal matter of animals (Allsop & Stickler, 1985; Kreader, 1995).

Kreader (1995) developed a PCR-hybridisation assay that targeted the 16S RNA gene sequence of *B. vulgatus, B. distasonis* and *B. thetaiotamicron*. Although results appeared promising, the PCR primers were not specific to *Bacteriodes* species of human origin. Depending on the primer set used, either 78% (*B. distasonis* and *B. thetaiotamicron*) or 67% (*B. vulgatus*) of the human faecal extracts contained high levels of target DNA. In comparison, 7 - 11% of the non-human extracts contained equivalent levels of target DNA, while an additional 12 - 20% contained target DNA at levels 100-1000 fold lower than detected in the human extracts. Of the primer sets, *B. vulgatus* was the least specific – up to 67% of domestic pets contained high levels of target DNA. Kreader (1995) concluded that using a combination of all three primer sets would distinguish human from non-human sources of faecal contamination (Kreader, 1995).

To date, two factors are paramount in limiting the application of *Bacteriodes* species as source specific indicators – cumbersome detection methodologies and poor survival rates in water. While PCR and hybridisation techniques may overcome the former limitation to some extent, the DNA must be sufficiently pure to eliminate PCR inhibitors. Secondly, the rate that *Bacteriodes* die-off in water is more rapid than that of either *E. coli* or *S. faecalis* (Allsop & Stickler, 1985; Fiksdal *et al.*, 1985; Kreader, 1998). Again, the latter limitation may be overcome using molecular techniques such as hybridisation with species-specific probes (Fiskdal *et al.,* 1985).

1.6.1.6 *Bacteriodes fragilis* Bacteriophage

Although *Bacteriodes* species appeared attractive surrogate indicators due to their high counts in human faeces, the bacteria did not survive well in the environment. Consequently, researchers investigated viruses (bacteriophage) infecting *B. fragilis*. Bacteriophages persist longer in the environment than the host bacterium and the structure, morphology, composition and size are similar to those of enteric viruses (Jofre *et al.,* 1986; Jofre *et al.,* 1989; Tartera & Jofre, 1987). As a result, bacteriophages may more accurately model disease risk of viral pathogens (Berg *et al.,* 1978; Grabow *et al.,* 1984; Grabow *et al.,* 1995). Studies on inactivation rates confirmed this, revealing that *B. fragilis* 40 phage exhibited similar decay rates to those of coliphage, poliovirus and rotovirus (Jofre *et al.,* 1986).

In the same way that the host bacterium, *Bacteriodes,* originates predominately from human faecal sources, so too does the corresponding phage. Evaluation of bacteriophage infecting *B. fragilis* 40 showed that all sample types polluted with sewage tested positive for the phage (Jofre *et al.,* 1986). In addition, counts reportedly outnumbered enteric viral counts by a factor greater than 10 (Jofre *et al.,* 1989). Following a survey optimising *B. fragilis* phage detection, Tartera & Jofre (1987) detected the phage in 10% of human faecal extracts when using the host with the highest detection efficiency — HSP40. This was supported by other researchers who reported that 13% of human faecal extracts contained the phage (Grabow *et al.,* 1995). In contrast, the phage was absent from non-human faecal samples and

samples which were not contaminated with faecal matter (Grabow *et al.,* 1995; Jofre *et al.,* 1986; Tartera *et al.,* 1987; Tartera & Jofre, 1987).

Recent research indicated that host selection is a determining factor for the specificity of phage detection from different faecal sources. For example, in a survey of different hosts, *B. fragilis* strains could be separated into two categories – those enabling detection from human sources (*e.g.,* HSP40) and those that enabled detection from both humans and animals (Puig *et al.,* 1997; Puig *et al.,* 1999).

The major drawback of utilising *B. fragilis* phage for differential purposes was the low frequencies that are encountered in humans and some waters — particularly coastal waters (Kai *et al.,* 1985; Sinton *et al.*, 1998). Phage counts were considerably lower than those of faecal coliforms or coliphage (Cornax *et al.,* 1991; Jagals *et al.,* 1995) and as a consequence some researchers failed to detect the phage (Armon & Kott, 1995; Donnison, personal communication, 1997; Jagals *et al.,* 1995). The inability to detect phage was also related to the host bacterium used. Consequently, enumeration necessitated protocols that were highly sensitive or that utilised sample concentration techniques, and that used a defined host bacterium (Jagals *et al.,* 1995). Such low frequencies effectively limited the value of using the organism for distinguishing human and animal faecal contamination. To date, detection using molecular techniques that may overcome this limitation to some extent, has not been reported.

1.6.1.7 F^+ RNA Coliphage Serotypes

F-specific (F^+) coliphage attach to bacterial 'F-pili' thought to be involved in the transfer of genetic material. Bacterial strains possessing F-pili are referred to as 'male' strains or F^+ strains (Donnison & Ross, 1995; Handzel *et al.,* 1993; Sinton & Finlay, 1996). The 'F-pili' are produced only when the host cells (*E. coli*, *Salmonella typhimurium*) are grown at temperatures above 30°C. In temperate countries natural waters rarely reach such temperatures, and consequently environmental multiplication of the phage is limited (Donnison & Ross, 1995; Havelaar & Pot-Hogeboom, 1988).

F^+ RNA phage are a subgroup of the F-specific (F^+) coliphage. F^+ RNA coliphage are small icosahedral phage containing single stranded RNA. They exhibit similar size, morphology and survival rates in fresh water to those of enteroviruses (Havelaar & Pot-Hogeboom, 1984; Havelaar *et al.,* 1993; IAWPRC, 1991; Sinton & Finlay, 1996). Consequently, F^+ RNA phages were suggested as useful models for the behaviour of human enteric viruses. F^+ RNA coliphage may be grouped into one of four serotypes (I-IV). Researchers found that serotypic classification also corresponded with faecal origin (Furuse *et al.,* 1981; Osawa *et al.,* 1981; Havelaar *et al.,* 1986). Serotype I coliphage originated from mammals other than humans or pigs, serotypes II and III originated from human faeces and sewage, and serotype IV indicated a mixture of both human and animal faecal contamination (Furuse, *et al.,* 1981; Havelaar *et al.,* 1986; Havelaar *et al.,* 1990; IAWPRC, 1991; Osawa *et al.,* 1981). Of the serotypes found in humans, type II coliphage were rarely found in faeces but were abundant in wastewater of human origin (Havelaar *et al.,* 1990). Pigs appeared to contain both serotypes I and II (Osawa *et al.,* 1981). Furuse *et al.* (1981) reported group I coliphage in raw sewage from treatment plants – although the plants also received effluents from slaughterhouses.

Because serotyping methodologies are expensive and laborious, routine typing of coliphage for environmental monitoring is unsuitable. The use of hybridisation techniques was investigated (Hsu *et al.,* 1995) resulting in 96.6% of F^+ RNA coliphage being assigned to the correct faecal source based on genotypic studies, compared with 99.5% using serotyping. These results indicated that hybridisation techniques might provide a suitable alternative to serotyping. Overall, however, F^+ RNA coliphage serotypes have not been comprehensively studied.

1.6.1.8 Summary of Differentiation Based on Species Composition

Protocols for differentiating sources of faecal contaminants, which are based on differences in the species composition from different faecal sources, either do not provide sufficient discrimination or, where differentiation is possible, are of restricted value due to limited survival or labour intensive methods. In addition, extensive epidemiological studies would be necessary to establish a correlation

between indicator frequencies and disease incidence. The scope of world-wide epidemiological studies required would prove immense, both economically and logistically. Hence several options for research are apparent. Either research should pursue indicator systems which provide substantial differentiation that may justify undertaking extensive epidemiological surveys, or investigations could be undertaken on current water quality indicators to assess if they could provide differentiation at the phenotypic or genotypic level.

1.6.2 Phenotypic Differentiation

1.6.2.1 Antibiotic Resistance

The advent of antimicrobial substances to control infection has unwittingly provided a new variety of emerging pathogens – those resistant to the very drugs which were designed for their demise. While antibiotic resistance of organisms has been a concern for many years, it is only recently that researchers have investigated a minor advantage of this – possible differentiation of human and animal faecal contamination. Although resistant strains may have emerged as a result of medicinal use of antibiotics, perhaps of far greater concern is the administration of these substances to animals as food additives for growth promotion and disease prevention. Antibiotic resistant *Salmonella* isolated from human infections correlate more closely with antibiotics used in animal feeds than those used for combating human infections (Angulo, 1999; Krumperman, 1983; Soothill & Mountford, 1999). The basis for investigating antibiotic resistant strains to differentiate sources of faecal pollution is that the antibiotics used to combat human infection are generally different from those used in animal populations. Hence the spectrum of resistance may differ in strains from different faecal sources. Based on this premise, researchers have examined the spectrum of antibiotic resistance amongst strains from two groups of indicator bacteria – the enterococci/faecal streptococci and faecal coliforms/*E. coli*.

1.6.2.1.1 Faecal streptococci/Enterococccci

Enterococci, a subset of the faecal streptococci group, are utilised as water quality indicators in estuarine or coastal ecosystems. Studies show differences in the

antibiotic resistance profiles of isolates from different faecal sources. Differentiation of human and animal isolates (ARCC[1] 95%) was indicated by discriminant analysis of resistance against four antibiotics at four concentrations when the isolates were pooled to two groups (human and animal) (Wiggins, 1996). These results were supported by Hagedorn *et al.* (1999) who used six antibiotics (ARCC 87%) and Harwood *et al.* (2000) using nine antibiotics (ARCCs of 75.5% for human and 72.4% for animal isolates). However, when isolate groupings were maintained at the level of individual sources the classification rates declined. Wiggins (1996) separated isolates into four faecal sources (human, cattle, poultry and wild) and obtained an ARCC of 84%. Harwood *et al.* (2000) and Hagedorn *et al* (1999) observed similar reductions in the rates of correct isolate classification. The incidence of false negatives varied significantly. Twenty one percent of known cattle isolates (Wiggins, 1996) and 30.5% of human isolates were mis-classified as originating from other faecal sources. In addition, 65% of chicken isolates were mis-classified; 20% were mistakenly assigned to human sources, 30% to wild faecal sources and the other 15% were distributed between cattle, dogs and swine sources (Harwood *et al.*, 2000).

Antibiotic selection appears to be an important factor in optimising classification of strains. Surprisingly, more is not necessarily better. Of five antibiotics tested (chlorotetracycline hydrochloride, halfuginone hydrobromide, oxytetracycline hydrochloride, salinomycin sodium, streptomycin sulphate), the best results were obtained using varying concentrations of just four antibiotics (*i.e.* excluding halfuginone hydrobromide) (Wiggins, 1996; Wiggins *et al.,* 1999). Harwood *et al.* (2000), however, obtained lower ARCCs when data from any of the nine antibiotics were omitted from the analysis. The less accurate classification obtained by Harwood *et al.* (2000) and Hagedorn *et al.* (1999) may have been attributable to either the combination of antibiotics used or greater strain diversity obtained from more comprehensive surveys. Although differentiation of faecal sources may appear feasible, enterococci are not recognised as suitable indicators in fresh water

[1] ARCC - Average rate of correct classification

environments because strains can originate from natural sources (*e.g.* insects, leaf decay) (Sinton *et al.,* 1993b). Having said that, it would be expected that the anti-microbial profiles of natural isolates would differ from those of either human or animal sources.

1.6.2.1.2 E. coli/*Faecal coliforms*

Although antibiotic resistance patterns of *E. coli* have been studied for over twenty years, little research has investigated their application for differentiating faecal sources. Early reports indicated that there was no obvious correlation between the incidence of antibiotic resistance and water source or level of contamination, irrespective of the antibiotics screened (Niemi *et al.,* 1983). In contrast, Krumperman (1983) reported that multiple antibiotic resistance (MAR) could be used to differentiate isolates originating from low-risk enteric *E. coli* sources and those obtained from high-risk sources, such as human, swine, commercial poultry farms, dairy cattle and associated rodent faecal matter. MAR indices from isolates originating from low-risk sources were typically <0.199 (*e.g.* grazing cattle, sheep, wild, rural rodents). In comparison, MAR indices from high-risk sources were >0.2.

The ability to differentiate faecal source using antibiotic resistance profiles was supported by Parveen *et al.* (1997), who showed that *E. coli* from point sources of contamination were more resistant and exhibited higher MAR indices than those from non-point sources. Isolates from human and animal faecal sources clustered amongst isolates from both point and non-point sources. Non-point sources are likely to be dominated by grazing cattle and sheep, whereas inputs from humans, swine, poultry farms and dairy cattle dominate point sources. Harwood *et al.* (2000) combined profiles of resistance to eight antibiotics and discriminant analysis. Using this protocol 69.3% of human isolates and 78.4% of pooled animal isolates were classified correctly. Isolates were further classified to individual sources as follows: human (54.2%), chicken (57.4%), cattle (54.5%), dogs (94.8%), swine (72.5%) and wild (50.6%). Unfortunately up to 50% of isolates from a faecal source were often mis-assigned.

Antibiotic resistance profiles of faecal streptococci or faecal coliforms appear to provide differentiation of faecal sources, especially when combined with discriminant analysis. Significant numbers of isolates, however, may be mis-classified. The technique requires additional manipulation of samples and greater technical expertise for data interpretation compared with standard methods for enumeration of *E. coli* or enterococci. Hence although the technique may be suitable for routine monitoring, it may be more applicable to analysis of special interest samples (*e.g.* prosecution cases, significant pollution events). Verification of the survival of antibiotic resistant strains may also be required as limited studies have indicated lower survival rates of resistant strains compared with non-resistant strains (Pettibone *et al.,* 1987). In addition, a large reference database must be established to enable classification of unknown strains.

1.6.2.2 Biochemical patterns

Biochemical fingerprinting of faecal coliforms (of which *E. coli* is typically a major component) has been used to compare isolates from different sources (Kühn *et al.,* 1991; Kühn *et al.,* 1997b). The authors used an array of biochemical assays in the form of the Phene Plate (PhP) to measure the rate of substrate utilisation. The system was originally developed to type strains either belonging to or related to *Enterobacteriaceae.* Previous authors have warned that typing systems often provide little discrimination of strains, as the main purpose is identification of isolates rather than assessment of genetic diversity (Crichton & Old, 1982).

Regardless, Kühn *et al.* (1997) found that suspected contaminating sources could be eliminated from the investigation on the basis of biochemical fingerprints. River isolates were found to harbour an extensive biochemical diversity, whereas the suspect contamination of pulp and paper effluent comprised isolates that exhibited limited diversity. Consequently the authors concluded that the pulp and paper mill effluent was not the principal source of downstream coliform contamination, but that several diffuse sources were responsible. Although biochemical typing may be useful, further research is required to ascertain whether the system provides sufficient strain discrimination to generate conclusive or useful results.

1.6.2.3 Summary of Phenotypic Differentiation

To date, of the phenotypic methods for differentiation of faecal origin, antibiotic resistance of indicator bacteria shows the most promise. In fact at times, the rate of successful classification appears remarkable (*e.g.* use of enterococci). Further research is required to optimise the method to increase the rate of correct classification. Such research must be undertaken using significant numbers of isolates of known faecal origin. Furthermore, research may be required on the survival of antibiotic resistant strains compared with the total *E. coli* population. Differentiation using either antibiotic resistance or biochemical profiles, may nevertheless be unsuitable for routine monitoring, because of the extensive time requirements for analysis and interpretation. Antibiotic resistance however may be suitable for defining the origin of "special interest samples", such as significant pollution events, blatant consent violation, or sites with consistently degenerating water quality.

1.6.3 Genotypic Differentiation

Presumably, because *E. coli* is recommended as one of most efficacious water quality indicators, much of the research investigating genotypic differentiation of faecal sources has targeted *E. coli.*

1.6.3.1 REP & ERIC primers

Recent years have seen the emergence of a variety of DNA fingerprinting techniques. The DNA of Enterobacteriaceae contains several repetitive chromosomal sequences: Repetitive Extragenic Palindromic (REP) sequences and Enterobacterial Repetitive Intergenic Consensus (ERIC) sequences. These techniques use PCR to amplify DNA between the repetitive elements to generate strain-specific DNA fingerprints (Dombek *et al.,* 2000; Lipman *et al.,* 1995). REP-PCR and ERIC-PCR have been used to investigate the epidemiology of *E. coli* in cattle with mastitis (Lipman *et al.,* 1995). Lipman *et al.* (1995) indicated that strain differentiation was possible using both methods, but that ERIC-PCR was more reliable than REP-PCR. The method was capable of differentiating strains exhibiting the same serotype, however

occasionally identical genotypes were shown to exhibit different serotypes. Therefore the authors recommended that a combination of serotyping and ERIC-PCR genotyping be used for epidemiological studies.

Other researchers utilised the same REP-PCR primers as Lipman *et al.* (1995), to generate fingerprints from *E. coli* isolates from animals and humans, in an attempt to differentiate sources of faecal pollution (Dombek *et al.*, 2000). Statistical analysis indicated that 100% of cattle and chicken isolates were correctly classified. Moreover between 78 – 90% of human, pig, sheep, goose and duck isolates were assigned to the correct groups. Consequently the results indicated that differentiation of faecal sources was possible and that the method may be an effective monitoring tool.

Several drawbacks are apparent for the application of the technique to routine environmental monitoring. The protocol in the first instance was very similar to standard methods for the enumeration of *E. coli* with a minor extension in analysis time due to PCR. The major drawback concerns data interpretation. Genomic fingerprints are complex and consequently visual examination is time consuming and unless systematically analysed, is subject to error. Data analysis must therefore be computer-assisted using gel analysis software, statistical analysis and comparison to a database of fingerprints. Nevertheless visual examination of the raw data is essential to minimise errors which may be generated by software analysis. All of these steps increase analysis time, cost and technical expertise required. The protocol apparently provides definition of faecal inputs, and as such may be useful for specific monitoring purposes such as supporting prosecution of blatant consent violations. Nevertheless an extensive reference database comprising the genomic fingerprints of isolates from various sources must be established by the monitoring agency prior to application of the method.

1.6.3.2 RAPD-PCR: Location of a Molecular Marker Diagnostic of Sewage

Randomly Amplified Polymorphic DNA, using the Polymerase Chain Reaction, (RAPD-PCR) utilises arbitrary primers to amplify random regions from the DNA of

an organism (Welsh & McClelland, 1990; Williams *et al.,* 1990). Genomic fingerprinting by RAPD-PCR has been used to locate a putative marker that correlated with ecological niche (Turner *et al.,* 1997a). RAPD-PCR profiles indicated distinctive differences between banding patterns obtained from "human" *E. coli* isolates (clinical and sewage derived), and those obtained from animal isolates (Turner, 1996). The animal profiles comprised bands over a wider size range, and included a number of prominent bands (250-500 bp) that were not evident in the profiles from human isolates (Turner, 1996).

These profile distinctions enabled Turner (1996) to identify a putative marker indicative of human origin. The marker, a 1.6 kb polymorphic fragment, was found in 60% of clinical *E. coli* isolates, 29% of sewage derived isolates and was present in 5% of bovine, ovine and equine isolates (Turner *et al.,* 1997a). Sequence analysis indicated that the putative marker was located downstream of the region encoding the glycine decarboxylase P-protein (gcvP). Although results were not as specific as was hoped, the relatively high occurrence in human isolates compared with animal isolates indicated that the marker might be useful for defining the origin of faecal contamination in the environment. Turner *et al.* (1997b) consequently examined the environmental application of the marker in a field study. A stream was selected that received effluent from a community treatment plant, but with no known upstream human sewage inputs. Marker positive isolates were not detected in the upstream samples. However, they were isolated from the effluent and downstream site, supporting the close association of the marker with human-derived *E. coli*. The conclusions from Turner's investigations (1996) supported the theory that host-related genetic variation within species may occur, and that these differences may be utilised to differentiate sources of faecal contamination.

To date there are no published data regarding the occurrence of the marker in non-clinical *E. coli* isolates. If the marker is solely associated with clinical isolates it may provide an accurate indication of disease risk associated with pathogenic *E. coli*. However, it would not indicate the presence of viral or protozoan disease risk, as would a more general indicator of faecal contamination.

1.6.3.3 Summary of Genotypic Differentiation

Genotypic differentiation appears to provide a feasible strategy for identifying the faecal origin of contaminants in aqueous environments. REP-PCR was capable of differentiating human and animal isolates with a high degree of accuracy. The profiles obtained can however be complex, and hence require computer-assisted gel analysis and statistical analysis of the fingerprints, to assign strains to faecal origin. As such REP-PCR provides a tool for the analysis of "special interest samples", but is unsuitable for routine environmental monitoring.

RAPD-PCR has been used to locate a marker for diagnosing human faecal contamination, which was developed into a method suitable for semi-routine monitoring. Application of fingerprinting methods for locating markers and then developing target-specific screening methods reduces the complexity of data interpretation and the analysis time. Such method development may facilitate routine or semi-routine monitoring.

1.7 *ESCHERICHIA COLI*

1.7.1 Introduction

Escherichia coli is a common constituent of the intestinal tract of warm-blooded animals. It is essentially a commensal organism in the animal host but some strains are pathogenic (Mason & Richardson, 1981; Hinton, 1985). Although *E. coli* represents less than 1% of the total faecal microbial population, this equates to 10^5-10^8 cells/gram in human faeces (Hartley *et al.*, 1977; Moore & Holdeman, 1974). *E. coli* also occurs in waterways, soil and sediment as a consequence of faecal contamination (Hartl & Dykhuizen, 1984) and are routinely monitored in New Zealand and overseas as an indicator of faecal contamination in fresh waters (Edberg *et al.*, 2000; Edberg *et al.*, 1997; Ministry for the Environment, 1999a).

1.7.2 *E. coli* as an Indicator

E. coli is the only member of the coliform group that is definitively associated with the faecal flora of humans and animals. Several advantages exist for using a marker present in *E. coli* for differential purposes. *Escherichia coli* is itself already recommended as the most efficacious indicator, particularly of fresh waters, for public health protection (APHA, 1995; Baudišová, 1997; Edberg *et al.*, 1997; Edberg *et al.*, 2000; Ministry for the Environment, 1999a). As such, standard methods for the direct detection of *E. coli* are readily available, sensitive, inexpensive and specific. In addition, detection methods are succinct and relatively easy to perform (Edberg *et al.*, 1997; Ministry for the Environment, 1999a; Ministry for the Environment, 1999b). Although the current water quality guidelines for fresh water are based on a limited number of epidemiological studies, a New Zealand research programme is currently underway to develop more robust guidelines for the recreational use of fresh water (Ministry for the Environment, 1999a).

Concerns have been raised that the survival of *E. coli* is significantly less than pathogens and hence may not be sufficiently indicative of viral disease risk (Berg *et al.*, 1978). Critics of *E. coli* have not, however, defined survival times appropriate for indicator organisms other than to say that survival should be greater than the pathogens that the indicator theoretically represents. The variety of pathogens present in water inherently confers much variation in their survival; therefore defining a single survival period is unrealistic (Edberg *et al.*, 2000). In addition, the survival of individual organisms is significantly affected by a variety of parameters, including temperature and predation (Serrano *et al.*, 1998) and sunlight inactivation (Davies-Colley *et al.*, 1994; Sinton *et al.*, 1999). Microbial survival is significantly enhanced when bacteria are sequestered by sediment particles (Davies *et al.*, 1995; Fish & Pettibone; 1995). Such physio-chemical parameters may additionally interact in synergistic manner, further confounding definition of an appropriate survival period.

1.7.3 Host Specificity in *E. coli*

The ability to differentiate sources of faecal contamination in the environment at the phenotypic or genotypic level implies host association of strains. Bacterial populations often consist of distinct clones that have adapted to specific environmental niches. Selective forces imposed on the organism by environmental factors maintain polymorphisms involved in clonal adaptation (Reeves, 1992).

The *E. coli* population in the faecal flora is not static and strain turnover in humans and animals has been observed (Hartley *et al.*, 1977; Sears *et al.*, 1950). According to Sears *et al.* (1950) "at any particular time the *E. coli* flora of the human intestinal tract consists of strains that persist over relatively long periods of time, accompanied at times by not more than three or four other strains that maintain a tenure of a few days or a few weeks only" (pg 300). These strains are known as residents and transients respectively (Guinee *et al,* 1972; Mason & Richardson, 1981; Hartl & Dykhuizen, 1984; Taylor, 1961). Subsequent researchers argued that resident and transient strains were difficult to define as only strains that dominated the faecal flora were analysed (Hartley *et al.*, 1977). Instead Hartley *et al.* (1977) coined the terms "majority resident" and "minority resident" if strains were repeatedly isolated from an individual, and utilised "transient" to describe strains obtained once or irregularly. Although majority resident strains have been shown to be displaced by episodes of diarrhoea (Mason & Richardson, 1981; Sears *et al.,* 1956) or antibiotic treatment (Petrocheilou *et al.,* 1979), attempts to establish new resident strains have been largely unsuccessful (Gage *et al.,* 1961; Sears *et al.*, 1950). The presence of "majority resident" *E. coli* strains in the human intestinal tract supports the theory of niche-specific selection of strains possessing adaptations that enable successful resistance to competition by other strains.

Early research utilising serotyping of *E. coli* reported that from over 14,000 isolates, 708 serotypes were observed. Of these, 520 originated from humans only, 130 from animals only and 58 occurred in both humans and animals (Bettelheim *et al.,* 1976). A later study compiling data from 15 years reported similar results (Bettelheim, 1978). Studies investigating the genetic structure of *E. coli* have predominantly

examined strains from human origin, and by comparison very limited analysis of animal isolates has been undertaken. The apparent host association of strains may simply be a function of the number of isolates analysed, such that host associations would disappear with a more extensive analysis of animal isolates. Alternatively, such host association of strains may in fact be "real". One of the difficulties with serotyping is the significant proportion of human and animal isolates that are difficult to serotype, in spite of the large number of antisera employed (Bettelheim *et al.*, 1976; Hartl & Dykhuizen, 1984).

Ochman *et al.* (1983) reported that although geographical clustering of strains was observed, no host association was apparent among humans and non-humans, carnivores and herbivores, or urinary tract isolates from infected versus healthy individuals. Instead, relatively few multilocus genotypes were found at much higher frequencies than would be expected. The authors concluded that the natural population of *E. coli* predominantly consisted of a mixture of independently evolving clones, some of which were widely distributed. The clonal theory, founded on results from multilocus enzyme electrophoresis analysis and biotyping, stated that the natural *E. coli* population was extremely diverse but organised into a limited number of clones (Ochman *et al.*, 1983; Milkman & Stoltzfus, 1988). Under this model the possibility of locating host association of strains would be unlikely (Turner, 1996).

Critchton & Old (1982) however warned that defining population structure based on a single typing method could lend to mis-classification of genotypically different isolates as single strains. Data have emerged in recent years that appear inconsistent with the clonal paradigm. Increasingly discriminatory methods have revealed that the *E. coli* population structure may be more complex than previously thought and may include extensive ecological subdivision of the population (Guttman, 1997). Souza *et al.* (1999) utilised a variety of techniques to examine the genetic structure of *E. coli* in wild hosts from several continents. Results indicated that genetic relationships amongst the strains were more strongly associated by origin and host than would be expected by chance, implying that host adaptation played an important role in the ecological structure of the population. These conclusions were strongly

supported by Dombek *et al.* (2000). Turner *et al.* (1997a) also reported host association of *E. coli*, as indicated by a marker which correlated with ecological niche.

1.8 LOCATING DIFFERENTIAL MARKERS

In summary, recent years have seen the escalation in research attempting to differentiate sources of faecal contamination in aqueous environments. To date, many of the protocols provide either insufficient differentiation or utilise laborious methods. One promising approach has targeted genomic fingerprinting (*e.g.* REP-PCR or RAPD-PCR). Research investigating REP-PCR analysis has predominantly emerged throughout the duration of the research described in this thesis, and hence these tools were not considered in initial research strategies. Moreover, the techniques do not lend themselves to routine monitoring of samples, because of the extended analysis requirements. Nevertheless, the literature on these methods has been included for completeness of the literature review.

Genotypic analysis using RAPD-PCR has indicated an ability to discriminate faecal sources of human origin (Turner, 1996). Hence, it was considered that genotypic comparison of isolates from various sources might provide the best chance at detecting differential host associations. Previous research on *E. coli* has indicated limited host-specific and geographical associations, however to date, few source-specific markers have been elucidated.

1.8.1 Genotyping Methods

Recent years have seen the emergence of molecular methods that are highly discriminatory; even closely related strains may be differentiated. Such differential power is necessary to undertake analysis similar to that described here. Amongst the emerging array of molecular methods developed for genomic characterisation of organisms, two methods have been shown to be useful for the strain-specific differentiation of isolates – Randomly Amplified Polymorphic DNA (RAPD) and

Amplified Fragment Length Polymorphism (AFLP). Both techniques are advantageous, in that a genomic fingerprint may be generated relatively easily and compared with fingerprints from other strains. Once suitable markers are located, target-specific primers may be developed to facilitate rapid screening of isolates.

1.8.1.1 Randomly Amplified Polymorphic DNA

Since RAPD-PCR (Welsh & McClelland, 1990; Williams *et al.*, 1990) utilises arbitrary primers to amplify random regions of DNA, no prior sequence knowledge is necessary and, theoretically at least, the technique may be applied to any organism (Power, 1996). RAPD-PCR is characterised by the use of short primers, typically ten nucleotides in length, and low stringency PCR, (*i.e.* low annealing temperatures are used during PCR). RAPD-PCR fragments are subsequently separated by gel electrophoresis and the profile obtained effectively constitutes a genomic fingerprint of the organism.

Amplification of random polymorphic regions of DNA has been used for a variety of differential purposes. Turner (1996) utilised the technique to locate a marker diagnostic of sewage contamination in clinical *E. coli* strains. Other researchers have since utilised the protocol for strain-specific identification of probiotic strains (Tilsala-Timisjärvi & Alatossava, 1998), differentiation of *Giardia muris* and *G. intestinalis* (Ionas *et al.*, 1997), and comparison of clinical and environmental strains of *Vibrio vulnificus* (Warner & Oliver, 1999). Hence the technique should be suitable for locating other faecal markers.

1.8.1.2 Amplified Fragment Length Polymorphisms

Amplified fragment length polymorphism (AFLP) is a novel DNA fingerprinting technique, capable of generating hundreds of reproducible markers. The relatively new technique confers the ability to simultaneously screen regions of DNA that are distributed throughout the genome. AFLP therefore enables high-resolution genotyping from the DNA of any organism without prior sequence information (Blears *et al.*, 1998; Mueller & Wolenbarger, 1999; Vos *et al.*, 1995).

Oligonucleotide adapters, which are ligated to restriction enzyme fragments, act as the binding site for PCR primers that selectively amplify a subset of the restriction fragment "population". PCR products are separated by gel electrophoresis and the visualised fragments create a genomic fingerprint. Polymorphic fragments are those that are present in one fingerprint but not another, and therefore may act as genotypic markers.

The protocol has been used to differentiate closely related strains that were previously impossible to resolve using alternative techniques (Aarts *et al.,* 1998; Desai *et al.,* 1998; Zhao *et al.,* 2000). In addition, a limited amount of research has been undertaken on the use of AFLPs for resolving strains from different sources (Dijkshoorn *et al,* 1996; Duim, Ang *et al.,* 2000; Janssen & Dijkshoorn, 1996). Although the technique has not been applied to differentiating faecal origin of strains, it may also be suitable for this purpose.

1.9 RESEARCH AIMS

The primary aim of the research described in this thesis was to locate a marker(s) indicative of agricultural faecal inputs in waterways. In addition, the feasibility of using such markers as environmental monitoring tools was also assessed. These research aims were achieved in the following manner:

- The use of Randomly Amplified Polymorphic DNA was assessed, for locating a marker(s) diagnostic of faecal inputs.

- The use of Amplified Fragment Length Polymorphisms was assessed, for locating a marker(s) diagnostic of faecal inputs.

- A method suitable for screening for the faecal marker(s) was developed.

- The specificity and prevalence of the faecal marker(s) was tested.

- The feasibility of using the marker in an environmental survey was examined.

2 Survey for Faecal Specific Markers – Randomly Amplified Polymorphic DNA

2.1 INTRODUCTION

2.1.1 RAPD-PCR Technique

Polymorphisms in genomic fingerprints may be utilised for the differentiation of closely related strains and may be detected using a variety of molecular techniques. One such technique is Randomly Amplified Polymorphic DNA analysis (RAPD) (Williams *et al.*, 1990) also known as Arbitrarily Primed PCR (AP-PCR) (Welsh & McClelland, 1990). RAPD analysis amplifies random regions of DNA using either single or multiple oligonucleotides of arbitrary sequence under low stringency PCR conditions and, theoretically, it is possible to apply the technique to any organism. RAPD-PCR products are separated by gel electrophoresis, effectively creating a genomic fingerprint of the organism. Genomic polymorphisms are fragments of DNA that are amplified from one strain but not from another (Welsh & McClelland, 1990; Welsh & McClelland, 1991; Williams, *et al.*, 1990).

RAPD analysis provides a number of advantages over other typing methods. The protocol is rapid, inexpensive and viable in most laboratories (Madden *et al.,* 1996; Power, 1996). As no prior sequence information is necessary, preliminary work such as the isolation of cloned DNA probes or nucleotide sequencing is also unnecessary (Williams *et al.,* 1990). The technique is applicable to a wide range of organisms and in fact a set of universal primers may be used for genotyping a wide variety species (Williams *et al.,* 1990). Moreover, the nature of RAPD analysis lends itself to automated genotypic profiling and hence enables more efficient typing than is achievable using other protocols such as pulse field gel electrophoresis (PFGE), southern blotting with insertion sequence (IS) probes or phage typing (Chatchaty *et al.*, 1994; Mueller & Wolfenbarger, 1999; Wang *et al.,* 1993). RAPD analysis has shown greater discriminatory power than multilocus enzyme electrophoresis

(MLEE), biotyping and RFLP analysis (Cave *et al.,* 1994; Wang *et al.,* 1993). The results obtained by RAPD-PCR are supported by those from other typing methods (Olmos *et al.,* 1998; Warner & Oliver, 1999).

RAPD methods have been used for a variety of applications, including strain-specific differentiation (Tilsala-Timisjarvi *et al.,* 1998), population and phylogenetic studies (Welsh & McClelland, 1991), discriminating swine-associated *Campylobacter coli* (Madden *et al.,* 1996), differentiating *Giardia muris* and *Giardia intestinalis* (Ionas *et al.,* 1997) and to trace the transport of *E. coli* through groundwater (Rothmaier *et al.,* 1997). In addition, Turner (1996) utilised RAPD analysis to identify a marker diagnostic of *Escherichia coli* strains from sewage. Consequently, RAPD analysis appears a suitable starting point for locating markers capable of differentiating the faecal origin of bacterial indicators.

2.1.2 Research Strategy

The research described in this chapter investigated the usefulness of RAPD analysis as a tool for locating genetic markers indicative of the faecal origin of bacterial indicators. To locate such markers, a genotypic fingerprint was generated using RAPD-PCR from *Escherichia coli* strains isolated from a variety of faecal sources. Dairy oxidation ponds, sewage treatment plants and agricultural run-offs are the predominant sources of faecal contamination in New Zealand waterways, hence isolates from these sources were examined. Once a marker(s) was located, suitable screening protocols may be developed – *e.g.* PCR primers derived from the makers may be used to amplify DNA which may discriminate between human and animal strains.

2.2 MATERIALS

2.2.1 Chemicals

Unless stated all chemicals used were analytical grade and were purchased from either BDH Laboratory Supplies (Dorset, UK) or Sigma Chemical Company (St Louis, MO, USA). Deionised water, herein described as Milli-Q water, was used for preparing all reagents and buffers.

2.2.2 Reagents

2.2.2.1 Buffer-saturated Phenol, pH 7.9

Phenol saturated with Tris-buffer, pH 7.9, was purchased from BDH Laboratory Supplies.

2.2.2.2 10% CTAB

10% CTAB comprised 10% hexadecyltrimethylammonium bromide in 0.7 M NaCl.

2.2.3 Enzymes, DNA Ladders and Primer Synthesis

Taq polymerase and 10 x *Taq* polymerase buffer (100 mM Tris-HCl, pH 8.3, containing 15 mM $MgCl_2$, 500 mM KCl,) were purchased from Roche Diagnostics (Auckland, NZ). 1 Kb Plus DNA Ladder^{TM} was obtained from Gibco BRL (Life Technologies, Rockville, MD, USA). Primers were synthesised by Life Technologies at 50 nmole scale and standard purity (Table 2.1).

2.2.4 Media

The following media were purchased from Difco Laboratories (Detroit, MI, USA); Bacto Peptone, mFc Agar, Nutrient Agar with MUG (4-methyl-umbelliferyl-β-D-glucuronide), Bacto-Agar and Luria Broth-Millers.

Table 2.1. Oligonucleotide primers

Function	Primer Name	Sequence 5' → 3'
E. coli verification	ECA75F	GGAAGAAGCTTGCTTCTTTGCTGAC
	ECR619R	AGCCCGGGGATTTCACATCTGACTTA
RAPD-PCR	GC30	TAATCACTGT
Fingerprinting	GC40	TAGTCACTGT
	GC50	TGGTCACTGT
	GC60	CGGTCACTGT
	GC70	CGGCCACTGT
	GC80	CGGCCACGGT
	GC90	CGGCCCCGGT
	GC100	CGGCCCCGGC

2.3 METHODS

2.3.1 Media Preparation

2.3.1.1 Sterilisation

Equipment, media and consumables were sterilised as necessary by autoclaving at 121°C for 15 min. Temperature sensitive buffers were sterilised by filtration through a sterile 0.2 μm filter.

2.3.1.2 Diluent

A solution of sterile 0.1% Bacto Peptone was sterilised and used as the diluent for all faecal samples.

2.3.1.3 Luria Broth-Millers; mFC Agar; Nutrient Agar with MUG

Luria Broth-Millers, mFC Agar and Nutrient Agar with MUG were prepared according to the manufacturer's instructions.

2.3.1.4 Luria Agar-Millers

Luria Agar was prepared as for the broth base, but with the addition of 15% Bacto-Agar.

2.3.2 Sample Collection

Human faecal specimens were obtained from seven healthy human subjects, aged between 20-50 years, who resided the Manawatu and Waikato regions. All subjects were from different households and were unrelated. Samples from aeration lagoons and oxidation ponds were collected near the pond outlets. Freshly excreted faecal specimens were collected from bovine and swine sources. The geographical origin of each of the samples is described in Table 2.2. All samples were collected in sterile pottles, stored on ice and processed, where possible, within six hours of sampling

Table 2.2. Geographical distribution of faecal samples

Origin	Sample Type	Geographical Location
Human	Stool	Manawatu, Waikato
	Sewage (Second Aeration Lagoon)	Manawatu, Taranaki
Animal	Bovine	Manawatu, Taranaki, Bay of Plenty
	Dairy (Secondary Oxidation Pond)	Manawatu, Taranaki
	Swine	Bay of Plenty
	Piggery (Primary Oxidation Pond)	Taranaki

2.3.3 Isolation of Faecal *Escherichia coli*

Faecal coliforms were enumerated by membrane filtration as described in Section 9222D of Standard Methods for the Examination of Water and Wastewater, 18th Edition (APHA, 1995). Fifty millilitres of sample were mixed by vortex in the presence of 1 g of 3 mm glass beads (Scientific Supplies Ltd., Auckland, NZ) for 2 min, to ensure that the sample was homogeneously mixed and to eliminate bacterial clumping. Serial dilutions were prepared from the sample using 0.1% peptone broth and filtered though sterile 0.2 μm gridded membranes (Gelman). Membranes were placed onto mFC agar and incubated at 44.5°C ($\pm$ 0.2°C) for 24 h, for the identification of faecal coliforms. After 24 h, membranes were transferred to Nutrient Agar with MUG and were incubated for a further 4 h at 35°C. *E. coli* isolates were identified as those colonies exhibiting blue fluorescence under long-

wave length ultraviolet light (366 nm) (Shadix & Rice, 1991). Isolates were streaked onto Luria Agar-Millers and incubated overnight at 37°C. Individual colonies were sub-cultured into 5 mL of Luria Broth-Millers, incubated at 37°C until the cultures reached mid log phase, then stored at –80°C in 15% glycerol. Typically, 15 to 25 faecal *E. coli* isolates were obtained from each source.

2.3.4 Isolates Used to Screen for Markers

Twenty *E.coli* isolates were chosen from a variety of faecal sources. Four isolates were selected from at least two different geographical locations and from each of the following sources – secondary sewage and dairy oxidation ponds, cattle stool samples and human stool samples. Two isolates from a primary piggery oxidation pond and two isolates from fresh pig manure were also analysed.

2.3.5 Genomic DNA Miniprep

Purified *E. coli* genomic DNA, prepared as described by Ausubel *et al.* (1994), was used as the template DNA for RAPD-PCR analysis. One and one half millilitres of an overnight culture was concentrated by centrifugation at 15,600 *x g* for 2 min. The supernatant was discarded and the pellet resuspended in 567 µL 1 x TE buffer (Appendix A.2). Thirty microlitres of 10% SDS solution and 3 µL 20 mg/mL proteinase K (Roche Diagnostics), were added to obtain a final concentration of 100 µg/mL proteinase K in 0.5% SDS. The sample was mixed by inversion then incubated for 1 h at 37°C to facilitate cell lysis. One hundred microliters of 5 M NaCl was added, followed by 80 µL 10% CTAB. The sample was mixed thoroughly after the addition of each reagent and incubated at 65°C for 10 min.

An equal volume of chloroform/isoamyl alcohol (24:1, v/v) was added, the sample was mixed by inversion, then centrifuged at 15,600 *x g* for 5 min. The supernatant was removed to a clean tube and another equal volume of phenol/chloroform/isoamyl alcohol (25:24:1, v/v) was added. The sample was mixed thoroughly, then centrifuged at 15,600 *x g* for 5 min. The supernatant was again transferred to a new

tube, with the addition of 0.6 vol. of isopropanol. The sample was mixed carefully until a "stringy-white" DNA precipitate formed. The precipitate was collected by centrifugation at 15,600 *x g* for 2 min, at room temperature. The DNA was washed with 70% ethanol, re-collected by centrifugation and dried at 37°C for 60 min. The DNA pellet was dissolved in 20 µL TE buffer (Appendix A.2) and stored at 4°C.

2.3.6 Agarose Gel Electrophoresis

The extracted DNA was assessed by electrophoreses on a 2% agarose gel. Gels were prepared with High Strength Analytical Grade Agarose (BioRad Laboratories, Hercules, CA, USA) in 1 x E-buffer (Appendix A.1). One microlitre of DNA was diluted with 2 µL loading buffer (Appendix A.3) and 4 µL Milli-Q water. High DNA Mass™ Ladder was loaded in adjacent lanes and gels were electrophoresed at 80V for 60-90 min. DNA was visualised by staining the gels in 100 µg/mL ethidium bromide and photographed under UV light using a 520LP filter on a Fluor-S™ MultiImager (BioRad Laboratories).

2.3.7 Verification of *E. coli*

Isolates were confirmed as *Escherichia coli* using target-specific PCR as described previously (Sabat *et al.*, 2000). PCR primers ECA75F and ECR619R (Table 2.1) targeted the 16S rRNA gene sequence of *E. coli*, enabling the amplification of a 570 bp product. The amplification reaction contained 2µL 10 x *Taq* polymerase reaction buffer, 0.5 mM $MgCl_2$, 200 µM each of dATP, dCTP, dGTP, dTTP (Roche Diagnostics), 40 pmole primer ECA75F, 40 pmole primer ECR619R, 1.5U DNA *Taq* polymerase and 1 µL genomic DNA template. The reaction volume was made up to 20µL with sterile Milli-Q water.

Conditions for PCR amplification were as follows: 1 cycle of 95°C for 30 sec; 40 cycles of 94°C for 30 sec and 72°C for 30 sec; 1 cycle of 94°C for 30 sec followed by a single cycle at 72°C for 5 min. Average ramp times were 1°C per second in a Perkin Elmer GeneAmp® PCR System 9700 (PE Biosystems, Foster City, CA,

USA). PCR products were separated by gel electrophoresis on a 1.6 % agarose gel, as described in Section 2.3.5. 1 Kb Plus DNA Ladder™ was run in adjacent lanes and used to determine product size.

2.3.8 RAPD-PCR Amplification

A set of RAPD-PCR primers (Williams *et al.,* 1990) that sequentially increased in % GC content (Table 2.1) were used for the analysis. The standard amplification reaction contained 2 μL 10 x *Taq* polymerase buffer, 250 μM each of dATP, dCTP, dGTP, dTTP, 1 pmole GC primer, 2.5 U DNA *Taq* polymerase and 1 μL genomic DNA template. The reaction volume was made up to 20 μL with sterile Milli-Q water.

PCR amplification conditions were as follows: 1 cycle of 95°C for 2 min, 35°C for 30 sec and 72°C for 30 sec; 30 cycles of 94°C for 30 sec, 35°C for 30 sec and 72°C for 30 sec; 1 cycle of 94°C for 30 sec, 35°C for 30 sec followed by a single cycle at 72°C for 5 min. Average ramp times were 1°C per second in a Perkin Elmer GeneAmp® PCR System 2400 (PE Biosystems). PCR products were separated by gel electrophoresis on 1.6 % agarose as described in Section 2.3.5. 1 Kb Plus DNA Ladder™ was run in adjacent lanes and used to determine product size.

2.4 RESULTS

2.4.1 Verification of *E. coli* Isolates

Twenty isolates were screened for markers indicative of faecal sources. These isolates were verified as *E. coli* using genus-specific primers, ECA75F and ECR619R, specific to the 16S rRNA sequence of *E. coli*. Figure 2.1 illustrates the *E. coli* genus specific-marker located at 544 base pairs (bp).

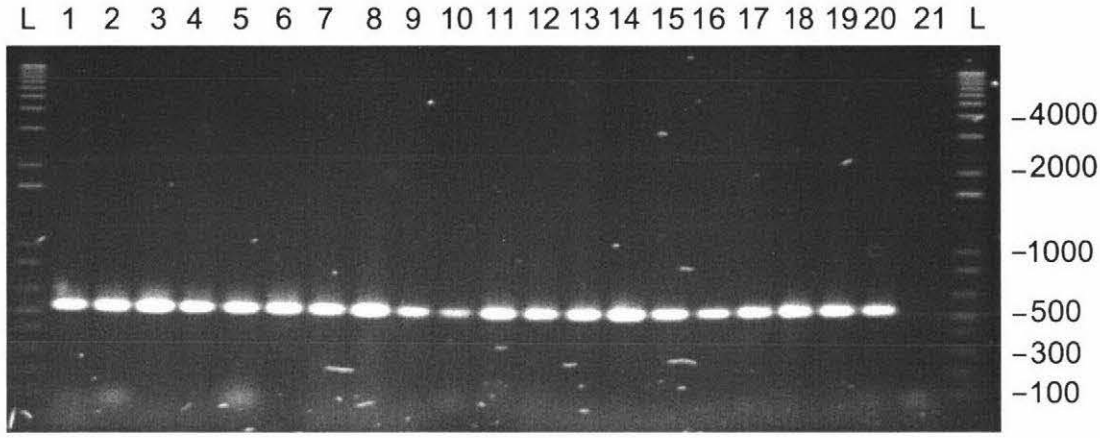


Figure 2.1. *E. coli* verification of isolates. Lane L, 1 Kb Plus DNA Ladder™; Lanes 1 - 4, sewage isolates; Lanes 5 – 8, human stool isolates; Lanes 9 – 12, dairy pond isolates; Lanes 13 – 16, cattle isolates; Lanes 17 - 18, piggery oxidation pond isolates; Lanes 19 – 20, pig isolates; Lane 21, negative control.

2.4.2 Genotypic Fingerprints Generated by RAPD-PCR

Electrophorectic analysis of RAPD-PCR products indicated highly polymorphic profiles (Figure 2.2). An overall increase in fragment generation was observed as the GC content of RAPD-PCR primers increased. Few, products (if any) were amplified using primer GC30, whereas primers GC90 and GC100 generated more complex fingerprints (data not shown). The majority of fragments produced by the RAPD-PCR primers screened were between 650 bp and 3 kilobases (Kb).

2.4.3 Putative Genus-Specific Markers

Two putative *E. coli* markers were identified. The first was a 660 bp fragment identified in genotypic fingerprints produced using the primer GC70 (Figure 2.2, Lanes 1-20). The second putative marker (1430 bp) was identified in the fingerprints produced using the primer GC90 (data not shown).

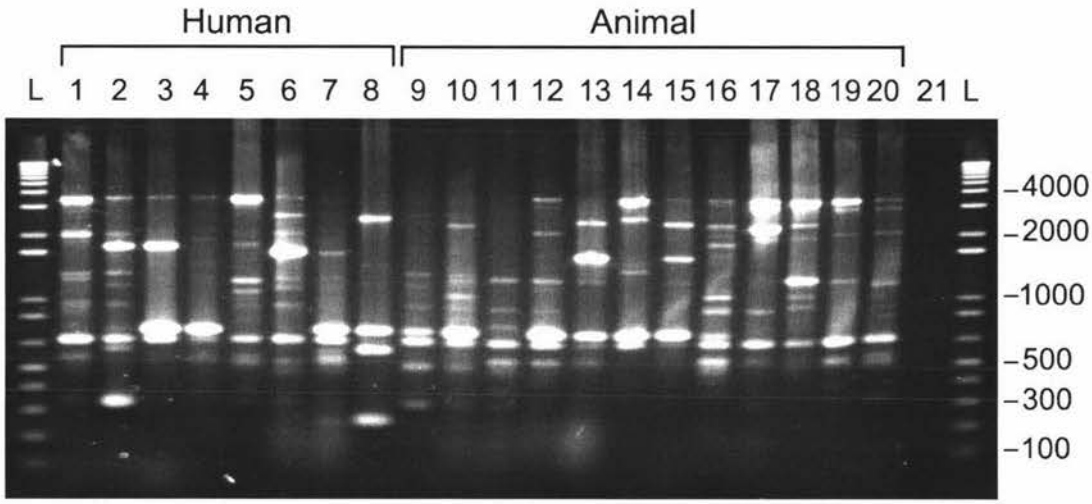


Figure 2.2. Genotypic fingerprints of faecal *E. coli* using primer GC70. Lane L, 1 Kb Plus DNA Ladder™; Lanes 1 - 4, sewage isolates; Lanes 5 – 8, human stool isolates; Lanes 9 – 12, dairy pond isolates; Lanes 13 – 16, cattle isolates; Lanes 17 - 18, piggery oxidation pond isolates; Lanes 19 – 20, pig isolates; Lane 21, negative control.

2.4.4 Diagnosis of Faecal Origin

Although putative genus-specific markers were apparent, singular bands did not correlated with particular faecal sources. The closest example was observed in fingerprints generated with primers GC40, GC50 or GC60. Fingerprints generated using these primers revealed a consistent fragment 1160 bp (Figure 2.3, Lanes 17-20), that was present in isolates from both piggery oxidation ponds and swine stool samples. Unfortunately, the fragment was also observed in two human isolates (Lanes 4–5) and a cattle isolate (Lanes 16), and hence did not provide sufficient discrimination of faecal origin. Although it was assumed that fragments of similar

size were the same product, it is possible that the 1160 bp fragment amplified from human isolates was a different product of similar size.

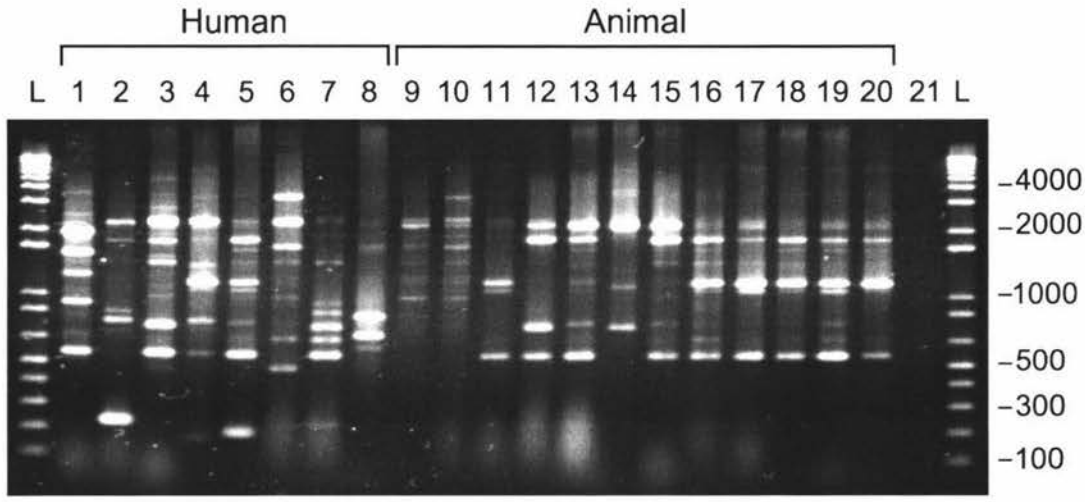


Figure 2.3. Genotypic fingerprints of faecal *E. coli* using primer GC60. Lane L, 1 Kb Plus DNA Ladder™; Lanes 1 - 4, sewage isolates; Lanes 5 – 8, human stool isolates; Lanes 9 – 12, dairy pond isolates; Lanes 13 – 16, cattle isolates; Lanes 17 - 18, piggery oxidation pond isolates; Lanes 19 – 20, pig isolates; Lane 21, negative control.

The genomic fingerprints obtained from human isolates (Lanes 1-8) were more diverse than those obtained from animal isolates (Lanes 9-20) and were generally unique for each isolate (Lanes 1-8). In contrast, animal (excluding dairy pond) isolates often exhibited one of two genotypic profiles (Lanes 13-15 and Lanes 16-20). The two genotypes were almost separated according faecal origin (cattle specimens versus swine isolates), but there was some crossover between the two groups (*e.g.* Lane 16 represented a cattle isolate, but exhibited a genotype that was similar to that of the swine isolates). The restricted banding profiles observed in animal (non-dairy pond) isolates was also apparent in one of the dairy pond isolates (Lane 12). Consequently, although the fingerprints were not entirely specific to faecal origin, generic clustering of isolates into human versus animal groups may be possible. The fingerprints obtained from piggery oxidation pond isolates were

similar to those from swine specimens, whereas the profiles from dairy oxidation pond isolates generally were not homologous to those from cattle specimen isolates.

2.5 DISCUSSION

2.5.1 Application of RAPD-PCR Analysis for Locating Diagnostic Markers

RAPD-PCR analysis provided excellent discrimination and rapid genotypic profiling of strains. The protocol was easy to perform and revealed putative genus-specific markers. Putative *E. coli* markers were not pursued, however, as several molecular methods are currently available for the diagnosis of *E. coli* and were utilised in this research (Sabat *et al.*, 2000). In spite of the discrimination afforded by RAPD-PCR, markers diagnostic of various faecal sources could not be identified.

Interestingly, the genotypic fingerprints of isolates originating from dairy oxidation ponds did not appear to exhibit similarity with those from cattle stool isolates. This was in contrast to the similarity observed in isolate fingerprints from pig stools and piggery oxidation ponds. Given that the dairy cattle stool isolates are the predominant (if not the sole) source of faecal input to dairy oxidation ponds, it was surprising that the genotypic fingerprints were not homologous. One factor contributing to this may have been that the piggery oxidation pond sample was taken from the outlet of the primary pond, whereas the dairy oxidation pond samples were taken from, or near the outlets of the secondary ponds. Therefore isolates from dairy oxidation ponds were subjected to longer exposure to external environmental stress (*e.g.* pH, solar radiation, temperature) (Davies-Colley *et al.*, 1999). As a result, the population structure of the dairy ponds may have changed significantly from that of the original faecal matter, because of differential die-off rates of the constituent strains.

2.5.1.1 Putative Swine Marker

The fragment that showed the most potential as a diagnostic marker was amplified from piggery and swine isolates. This putative "swine" marker was however

presumed identical to a fragment that was also present in human isolates and a single cattle isolate, and hence was not sufficiently diagnostic of faecal origin. The assumption that fragments of the same size were identical may or may not be correct. One way to clarify this supposition would be to sequence the fragment from animal isolates, design specific PCR primers and then screen isolates from various sources to ascertain the specificity of the marker. The putative swine marker was cloned and sequenced (Chapter Four), however limited investigation failed to obtain quality sequence information, presumably a result of the large size of the fragment. In addition, isolates in this study were obtained from a single sample and therefore were insufficiently representative of piggery faecal sources. Moreover, AFLP analysis was being concurrently trialed and indicated more suitable research targets. Consequently, this putative marker was not investigated further.

2.5.1.2 Differentiating Human and Animal Isolates

As no single marker was observed that correlated with strain niche, an alternative approach may be taken. The two genotypes observed in animal strains (excluding dairy pond isolates) might be useful for assigning strains to human versus non-human faecal origin. As a result, strains could be clustered on the basis of the entire genomic fingerprint rather than a single polymorphism. Various researchers have utilised comparative gel software and statistical analysis of genomic fingerprints to assign strains to closely related groups. Dombek *et al.* (2000) reportedly assigned 100% of cattle isolates and between 78 and 90% of human isolates to the correct faecal origin using a combination of REP-PCR and a Jaccard band-matching algorithm. A similar approach has also been applied to fingerprints generated by Amplified Fragment Length Polymorphism (AFLP) analysis (Crabill *et al.*, 1998; Zhao *et al.*, 2000).

Combining RAPD fingerprints and statistical analysis may be applied in a number of ways to discriminate isolate origin. First, the approach may be applied to primer combinations that exhibit restricted profiles for animal isolates and diverse profiles for human isolates (Figure 2.3), or *vice versa*, thereby providing generic

discrimination (animal versus human). Second, the approach may be applied to more complex fingerprints that harbour a combination of polymorphisms, which may be used to distinguish faecal origin. Third, it may be necessary to use several fingerprints from different primer combinations for each strain to enable sufficient discrimination. The latter two methods may provide identification of faecal origin at the source level (such as cattle, swine, human, *etc.*) rather than the generic "human versus non-human" discrimination. Methods utilising statistical analysis of fingerprints for discrimination would be more robust for assigning strains to the correct faecal origin as classification relies on numerous polymorphisms rather than a singular polymorphism.

Although the RAPD technique is a comparatively rapid genotyping protocol and is suitable for locating markers from a limited number of isolates, it is not suitable for routine environmental monitoring of large numbers of isolates. Nevertheless the technique may be suitable for monitoring "special interest samples" (*e.g.* prosecution cases; blatant non-compliance of consents; sites indicating significant faecal inputs of unknown origin). Results however, must be compared with a reference database comprising genotypic fingerprints of isolates from known sources.

2.5.2 Conclusions

RAPD-PCR analysis was investigated for locating markers indicative of various sources of faecal contamination. Once a marker was identified, the intention was to adapt alternative detection protocols (such as PCR amplification) to facilitate detection of the marker in a manner that is more conducive to screening large numbers of isolates. RAPD-PCR provided excellent discrimination of strains and was relatively rapid as a genotyping protocol where comparatively few strains are analysed. Nevertheless, a single marker that definitively correlated with faecal origin was not observed. Human genotypes were typically heterogeneous, whereas animal (non-dairy pond) isolates exhibited a number of common bands. The difference in genetic diversity of strains from various niches may provide suitable generic discrimination of faecal origin when combined with statistical techniques such as

discriminant analysis. Screening of environmental isolates by RAPD-PCR is unlikely to provide routine assessment of faecal origin.

2.5.3 Future Research

RAPD analysis did not indicate any putative faecal source markers over the limited period of this research. The following research strategies became apparent at the latter stages of the project.

1. RAPD-PCR fingerprints could be screened on polyacylamide gels (PAGE) using silver stain detection. This approach would enable better resolution and sensitivity of fragment detection, increasing the number of bands observed and the opportunity to detect polymorphic markers.

2. A greater number of RAPD-PCR primer combinations could also be screened.

3. RAPD-PCR primers could be used in combination to generate genomic fingerprints. Fewer than half of the fragments generated using primer combinations are present in fingerprints produced using the individual primers (Welsh & McClelland, 1991). Hence, this approach should enable generation of fragments that are not detected using individual primers.

4. Although the putative swine "marker" was not exclusive to isolates originating from pigs, it may not have been identical to the fragment observed in the human isolate. Therefore, the fragment could be sequenced from both swine and human isolates to verify if the fragments from both sources were homologous.

5. Since RAPD-PCR did not locate any definitive markers, alternative molecular fingerprinting techniques could be investigated. "Amplified Fragment Length Polymorphism" (AFLP) is a novel technique which is highly discriminatory

and has been used for sub-species differentiation of strains. This research is described in Chapter Three.

3 Identification of Faecal - Specific Markers – Amplified Fragment Length Polymorphism

3.1 INTRODUCTION

3.1.1 Amplified Fragment Length Polymorphism Analysis

Since RAPD-PCR fingerprinting (Chapter 2) did not clearly identify markers indicative of faecal origin, an alternative approach, Amplified Fragment Length Polymorphism, (AFLP) was investigated.

This relatively new technique confers the ability to screen regions of DNA that are distributed throughout the genome. The PCR-based technique simultaneously generates hundreds of reproducible markers, enabling high-resolution genotyping from the DNA of any organism without prior sequence information (Blears *et al.*, 1998; Mueller & Wolfenbarger, 1999; Vos *et al.*, 1995).

Essentially, total genomic DNA is digested with two restriction enzymes. Double stranded oligonucleotide adapters are ligated to the restriction fragments and the combination of the ligated adapter and restriction half site create the binding site for PCR primers. The Polymerase Chain Reaction is used to selectively replicate complementary restriction fragments to levels that may be visualised. The presence or absence of DNA fragments following electrophoretic separation indicates polymorphisms.

The technique is flexible and may be manipulated to control the complexity of the genomic fingerprint obtained. Restriction enzyme selection, and primer composition and length, may all be utilised to control the complexity of the genotypic

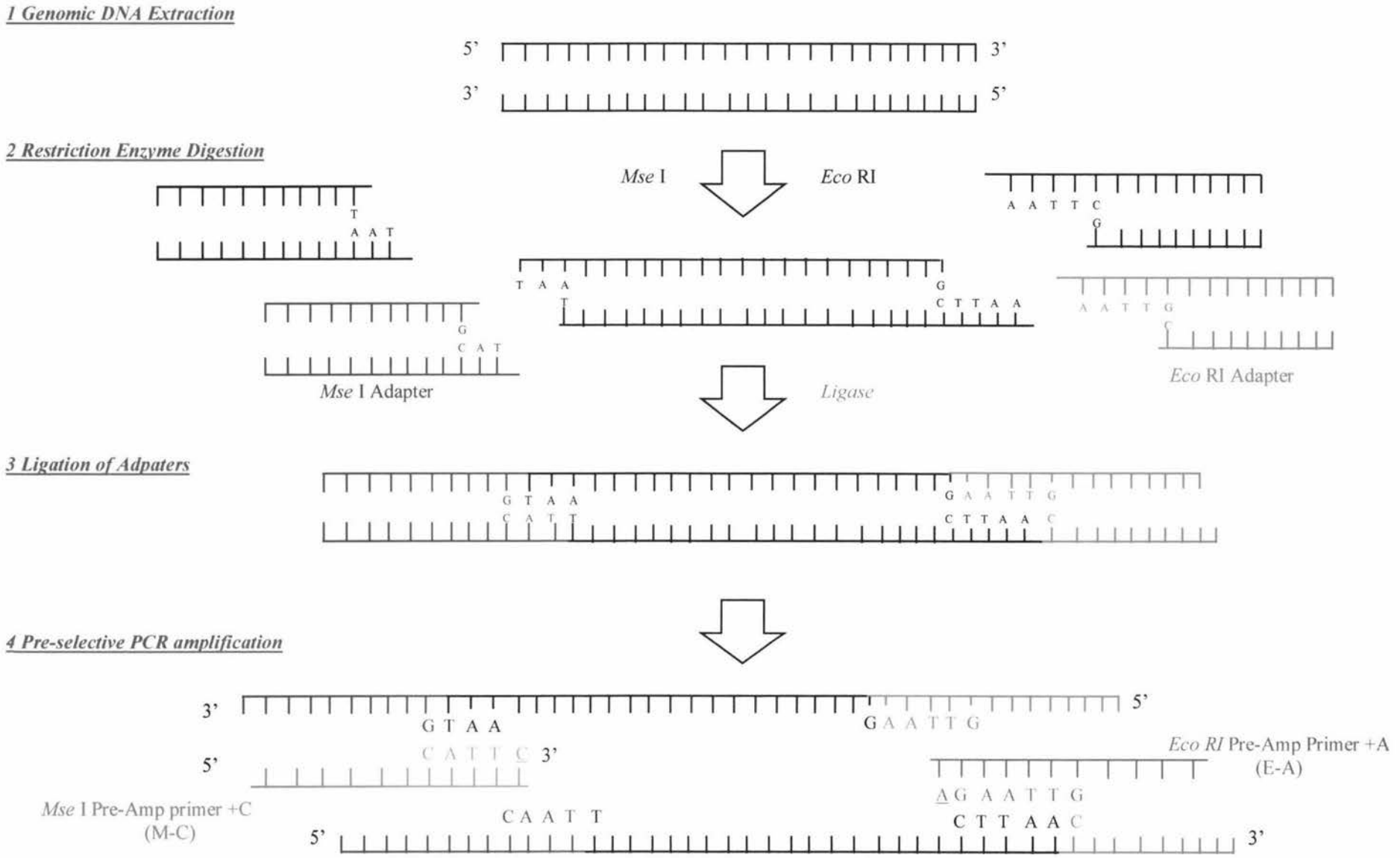


Figure 3.1. AFLP Schematic

(Mueller & Wolfenbarger, 1999)

fingerprints (Blears *et al.*, 1998). Typically, a combination of a rare and a frequent cutter restriction endonuclease are utilised. (Blears *et al.*, 1998). Janssen *et al.*, (1995) reported that *Mse* I/*Eco* RI was most suitable for analysis of low G+C content bacterial genomes whereas *Apa* I/*Taq* I were most suited for high G+C content genomes.

The AFLP PCR primers may contain additional selective nucleotides at the 3' end to reduce the complexity of fingerprints — typically between zero and three nucleotides. The addition of each selective nucleotide to the primer typically results in a four-fold reduction of fragments. The fragments obtained using increasingly selectively AFLP primers are typically a subset of the fingerprint obtained using a less selective primer (Vos *et al.*, 1995). In addition, a two-stage PCR protocol may be used to simplify genomic fingerprints, depending on the complexity of the genome being analysed — *i.e.* pre-selective PCR followed by selective PCR. Pre-selective amplification (Pre-Amp) was developed for complex genomes (10^8-10^9 bp) such as *E. coli* (2.3x 10^9-3.0x10^9 bp) (Hartl & Dykhuizen, 1984). Pre-Amp primers typically contain a selective overhang (either zero or one nucleotides), whereas the selective PCR primers contain between one and three selective nucleotides. The first nucleotide of the selective PCR primer is therefore identical to that of the Pre-Amp PCR primer (*e.g.* Pre-Amp PCR with E-A/M-C[1] may be followed by selective amplification using E-ATT/M-CTC).

The AFLP technique has a number of advantages over other genomic fingerprinting techniques. In the first instance, the protocol enables analysis of the entire genome, unlike techniques such as ribotyping or PCR-RFLP analysis of rDNA (Janssen *et al.*, 1996). Additionally, the protocol can generate genetic markers that are virtually free from artefacts (Mueller & Wolfenbarger, 1998). The latter is a consequence of the preciseness with which the AFLP primers match the target site. Although primer

[1] E-A represents the AFLP primer that is complementary to the *Eco* RI restriction half site plus the ligated adapter and the selective nucleotide overhang — adenine; M-C represents the AFLP primer complementary to the *Mse* I restriction half site plus the ligated adapter and the nucleotide overhang — cytosine.

mis-match may occur between the selective extension and the template, initiation of DNA synthesis will be inefficient or non-existent due to the high stringency PCR (Janssen *et al.*, 1996; Vos *et al.*, 1995).

Of foremost importance, the AFLP protocol enables extremely reproducible analyses. When high stringency PCR conditions are used the percentage error is reportedly less than 2%, including mis-priming and scoring errors (Mueller & Wolfenbarger, 1999). Numerous studies have shown that the results obtained with AFLP analysis were in agreement with those from alternative typing methods, but with superior discrimination (Arnold *et al.*, 1999; Janssen *et al.*, 1996; Janssen & Dijkshoorn, 1996). In addition, AFLP screening is cost effective, relatively rapid and suitable for automated analysis of genomes, which enables higher throughput than other molecular typing methods (Arnold *et al.*, 1999; Desai *et al.*, 1998; Koelman *et al.*, 1998). Due to the vast complexity of fingerprints that may be generated however, data must be systematically analysed to enable useful data capture and interpretation. The level of complexity should be optimised, based on the research target and available equipment. However, insufficient discrimination may lead to "clonal clustering" of genetically diverse strains.

Initially, use of the AFLP technique was dominated by plant genetic investigations. However recent years have seen increased application in bacterial fields. AFLPs have been utilised in a variety of applications including construction of high resolution genetic maps (Duim *et al.*, 1999), identification of clonal strains (Kühn *et al.*, 1997c), inference of phylogenetic relationships (Arnold *et al.*, 1999) and to aid epidemiological typing (Gibson *et al.* 1998; Janssen & Dijkshoorn, 1996; Valsangiacomo *et al.*, 1995). The protocol has also been used to differentiate closely related strains that were impossible to resolve *via* alternative techniques (Aarts *et al.*, 1998; Desai *et al.*, 1998; Zhao *et al.*, 2000). Limited research has also investigated the use of AFLPs for resolving strains from different sources. Kühn *et al.* (1997c) resolved human and environmental *Aeromonas* isolates, using a combination of AFLP analysis and FAME (fatty acid methyl ester). Differences were not, however, apparent between isolates originating from humans with diarrhoea and healthy

controls (Kühn *et al.,* 1997a). AFLP analyses have also successfully clustered *Acinetobacter* outbreak strains from sporadic isolates (Dijkshoorn *et al.,* 1996). Conversely *Campylobacter* strains from humans and chickens could not be differentiated (Duim *et al.*, 1999; Duim *et al.,* 2000).

3.1.2 Research Strategy

The research described in this chapter utilised AFLP analysis to locate markers indicative of various sources of faecal *E. coli* strains. Once a marker(s) was located, the intention was to develop a method suitable for rapid screening of isolates (*e.g.* PCR amplification). To locate diagnostic markers, a genotypic fingerprint was generated from strains from a variety of faecal sources. Effluents from dairy and piggery oxidation ponds, sewage treatment plants and non-point agricultural contaminants, are the dominant faecal inputs in New Zealand surface waters (Horizons Manawatu-Wanganui, 1999; Obiri-Danso & Jones, 1999). Hence these sources were targeted in the research.

3.2 MATERIALS

3.2.1 Chemicals

Unless stated all chemicals used were analytical grade and were purchased from either BDH Laboratory Supplies (Dorset, UK) or Sigma Chemical Company (St Louis, MO, USA). Deionised water, herein described as Milli-Q, was used for preparing all reagents and buffers.

3.2.2 Reagents

3.2.2.1 5 x Restriction Enzyme Buffer

5 x restriction enzyme buffer comprised of 50 mM Tris-HCl pH 7.5, 50 mM magnesium acetate and 250mM potassium acetate.

3.2.2.2 1x TBE Buffer

1x TBE comprised 89 mM Tris base, 89 mM boric acid, 1 mM EDTA pH 8.0.

3.2.2.3 Silver Staining Solution

The silver staining solution comprised 10 mg/L silver nitrate and 0.056% v/v formaldehyde.

3.2.3 Enzymes, DNA Ladders, Primer Synthesis, Restriction Endonucleases

Unless stated, *Taq* polymerase and 10 x *Taq* polymerase buffer (100 mM Tris-HCl, pH 8.3, containing 20 mM $MgCl_2$ and 500 mM KCl) were purchased from Roche Diagnostics (Auckland, NZ). DNA ladders was obtained from Gibco BRL (Life Technologies, Rockville, MD, USA). Primers and oligonucleotide adapters were synthesised by Life Technologies at 50 nmole scale and standard purity (Tables 3.1, 3.2 and 3.3). *Eco* RI was purchased from Gibco BRL and *Mse* I was purchased from New England Biolab *Inc.*, (Beverly, MA, USA).

Table 3.1. AFLP adapter sequences

Function	Name	Abbreviation	Sequence (5' → 3')
Adapter	*Eco* RI Adapter	*Eco* RI-1	CTCGTAGACTGCGTACC
		Eco RI-2	AATTGGTACGCAGTCTAC
	Mse I Adapter	*Mse* I-1	GACGATGAGTCCTGAG
		Mse I-2	TACTCAGGACTCAT

3.3 METHODS

AFLP analysis (Vos *et al.*, 1995), was used to develop a genomic fingerprint of faecal *E. coli* isolates from various sources. Prior to screening isolates with AFLP primers, conditions were optimised as described in Appendix B.

3.3.1 Sterilisation

Equipment, media and consumables were sterilised as necessary by autoclaving at 121°C for 15 min. Heat sensitive buffers were sterilised by filtration through a sterile 0.2 μm filter.

3.3.2 Sample Collection and Isolation of *E. coli* Strains

Twenty *E. coli* isolates were chosen from a variety of faecal sources (Sections 2.3.2 and 2.3.3). Four isolates were selected from at least two geographical locations, from the following sources – secondary sewage and dairy oxidation ponds, cattle and human stool samples. Two isolates from a piggery oxidation pond and two from fresh pig manure were also analysed. Isolate numbers were limited to facilitate concurrent screening of several faecal sources.

3.3.3 AFLP Analysis

3.3.3.1 Restriction Enzyme Digestion

Genomic DNA was prepared as described in Sections 2.3.4 and 2.3.5. DNA was digested with the following restriction enzymes: *Mse* I and *Eco* RI. Four units *Mse* I, 2 U *Eco* RI and 6 μL 5 x restriction enzyme buffer were added to 3 μL genomic DNA. The total volume was made up to 30 μL with sterile Milli-Q water and incubated at 37°C for 60 min. DNA digestion was confirmed by gel electrophoresis as described in Section 3.3.3.5.1.

3.3.3.2 Ligation of Oligonucleotide Adapters

Mse I and *Eco* RI adapters (Table 3.1) were prepared as described in Appendix A.5 and ligated to the digested DNA using the following protocol. Five pmoles of *Eco* RI adapter, 50 pmol *Mse* I adapter, 1 U T4 ligase, 3 μL 10 x T4 ligase buffer and 5 μL digested DNA were diluted to a final volume of 30 μL and incubated overnight at 4°C.

3.3.3.3 Pre-selective PCR Amplification

A variety of AFLP pre-selective and selective primers (Table 3.2 and 3.4) were used to analyse *E. coli* from various faecal sources. Pre-selective amplification using the Polymerase Chain Reaction (PCR) was performed as follows. "PreAmp" primers containing either zero or one additional selective nucleotide (Table 3.2) were used to amplify selected restriction fragments away from the total restriction fragment "population". Each PCR reaction contained 1 x *Taq* polymerase buffer, 250 μM of each dNTP (dATP, dCTP, dGTP, dTTP) (Roche Diagnostics), 10 pmole of each pre-selective primer (E-PreAmp, M-PreAmp), 1U *Taq* polymerase and 1 μL ligated DNA. Amplification conditions are outlined in Table 3.3. Average ramp times were 1°C per second in either a Perkin Elmer GeneAmp® PCR System 2400 or Eppendorf Mastercycler®. Products from the pre-selective amplification were stored at –80°C. Negative controls were run with all experiments, in which the DNA template was substituted with water.

Table 3.2. Pre-selective primers for AFLP amplification

PreAmp Primer	Abbreviation	Sequence
Eco+0[1]	E-0	GACTGCGTACCAATTC
Eco+A[2]	E-A	GACTGCGTACCAATTC+**A**
Eco+G[3]	E-G	GACTGCGTACCAATTC+**G**
Eco+T[4]	E-T	GACTGCGTACCAATTC+**T**
Mse+A	M-A	GATGAGTCCTGAGTAA+**A**
Mse+C[5]	M-C	GATGAGTCCTGAGTAA+**C**

[1] AFLP core primer sequence plus no selective nucleotide overhang; [2]selective overhang is adenine; [3] selective overhang is guanine; [4] selective overhang is thymine; [5] selective overhang is cytosine.

3.3.3.4 Selective PCR Amplification

A second amplification step, selective PCR amplification, was used to further reduce the complexity of the genomic fingerprint. PCR products generated by the pre-selective PCR step were subsequently used as the template DNA in selective PCR amplification. Selective amplification utilised primers containing either two or three selective nucleotide overhangs (Table 3.4). Consequently, the core sequence and the

first nucleotide overhang of the selective primer were identical to those used for the pre-selective primer amplification step. PCR reagent concentrations were similar to those used for pre-selective amplification, substituting the pre-selective primers for the corresponding selective primers and the ligated DNA was substituted with the Pre-Amp PCR product. Amplification conditions are described in Table 3.3.

Table 3.3. AFLP-PCR amplification conditions

PCR Protocol	No. of Cycles	Denaturation	Annealing	Extension
Pre-Amp	20 cycles	94°C; 30 sec	56°C; 60 sec	72°C; 60 sec
Selective	1 cycle	94°C; 2 min		
	6 cycles	94°C; 30 sec	65°C; 30 sec	72°C; 60 sec
	6 cycles	94°C; 30 sec	60°C; 30 sec	72°C; 60 sec
	24 cycles	94°C; 30 sec	56°C; 30 sec	72°C; 60 sec
	1 cycle			72°C; 5 min

Table 3.4. Selective primer combinations for AFLP amplification

	E-A[1]	E-G[1]	E-T[1]	E-AG[2]	E-AT[2]	E-AC[2]	E-GT[3]	E-GC[3]	E-GA[3]	E-AGC[2]	E-ATT[2]
M-C[4]	✓										
M-A[4]		✓		✓	✓	✓	✓	✓	✓		
M-AG[5]				✓	✓	✓	✓	✓	✓		
M-AT[5]				✓	✓	✓	✓	✓	✓		
M-AC[5]				✓	✓	✓	✓	✓	✓		
M-CG[6]				✓	✓	✓	✓	✓	✓		
M-CT[6]				✓	✓	✓	✓	✓	✓	✓	
M-CA[6]				✓	✓	✓	✓	✓	✓		
M-CTC[6]			✓							✓	✓
M-CTA[6]			✓							✓	✓

PreAmp primers used: [1]*Eco*+0 [2]*Eco*+A; [3]*Eco*+G; [4]*Mse*+0;[5]*Mse*+A; [6]*Mse*+C.
Eco RI primer and *Mse* I primer are abbreviated to E-N and M-N respectively, where N is the selective nucleotide overhang.
✓ indicates primer combinations tested.

3.3.3.5 Resolution of AFLP Products

Selective PCR products were separated initially on a 5% denaturing urea/polyacrylamide gel as described below. However to facilitate higher throughput screening, 2% agarose gels were subsequently used to screen for putative markers.

3.3.3.5.1 Agarose Gel Electrophoresis

AFLP products were separated by gel electrophoresis on a 2% agarose gel. Gels were prepared with High Strength Analytical Grade Agarose (BioRad Laboratories, Hercules, CA, USA) in 1 x E-Buffer (Appendix A.1). DNA was heated at 65°C for 10 min and then diluted with loading buffer (Appendix A.3). 1 Kb Plus DNA Ladder™ was loaded in adjacent lanes to determine product size and gels were run at 80 V for 75 min. DNA was stained the gels in 100 µg/mL ethidium bromide. Gels were photographed under UV light using a 520LP filter, on a Fluor-S™ MultiImager (BioRad Laboratories).

3.3.3.5.2 Denaturing urea/polyacrylamide gels

Denaturing urea/polyacrylamide gels were run using a Model S2 sequencing rig (Life Technologies) (gel size 31.0 x 38.5cm x 0.4 mm) and were prepared as follows. The long glass plate was coated with Rain-X (Unelko Corp., Scottsdale, AZ, USA) according to the manufacturer's recommendations. The short plate was coated with Bind-Silane (Amersham Pharmacia Biotech, Auckland, NZ) – 1 µL Bind-Silane, 2 mL 95% ethanol and 10 µL glacial acetic acid were spread onto the plate, wiped off and allowed to dry for 4 min. Excess Bind-Silane was removed with three washes of 2 mL 95% ethanol.

The acrylamide gel comprised 7 M urea, 5% Long Ranger® Gel Solution (FMC BioProducts, Rockland, ME, USA) and 1 x TBE, in a total volume of 70 mL. Immediately prior to pouring the gel, 350 µL 10% ammonium persulphate and 39 µL TEMED (N,N,N',N'-tetramethylethylenediamine) (BioRad Laboratories) were added. The gel was left to polymerize overnight. Excess acrylamide was removed prior to electrophoresis and the gel was pre-warmed at 45 W for 30 min. Four microlitres of

each selective AFLP-PCR product was mixed with 2 μL formamide loading dye (Appendix A.5). Molecular mass ladders were prepared by diluting 100 bp DNA Ladder™ with formamide loading dye to a final concentration of 1 μg/μL. Samples and standards were heat denatured at 95°C for 4 min. Polyacrylamide gels were run at 45 W until the bromophenol blue dye front was approximately 6 cm from the end of the gel.

PCR products were visualised by silver staining. The gel was soaked with gentle agitation, in 10% acetic acid for 4 hours to remove the urea. Acetic acid was subsequently removed using three two-min washes of Milli-Q water. The gel was transferred to the silver staining solution and agitated gently for 1 hour. Excess stain was removed with a brief, chilled (4°C) Milli-Q water rinse. The gel was transferred to chilled developer solution that was prepared as follows. Immediately prior to developing the gel 6.5 mL formaldehyde and 800 μL 10 mg/mL sodium thiosulphate were added to 0.1% sodium carbonate solution. The gel was developed until bands began to appear, transferred to fresh developer, and then agitated until the bands were clearly distinguished. Chilled 10% acetic acid was used to stop excess background development.

3.3.4 AFLP Reproducibility

To assess the reproducibility of the AFLP technique, quintuplicate genomic DNA extractions were undertaken from a single isolate and were analysed using the selective AFLP primers, E-AG and M-AG. AFLP-PCR products were separated on 2% agarose and the profiles obtained were compared for similarity.

3.4 RESULTS

3.4.1 Reproducibility of AFLP Genomic Fingerprints

Replicate AFLP fingerprints on quintuplicate DNA were identical for all analyses (data not shown).

3.4.2 Resolution of Genomic Fingerprints

AFLP analysis of *E. coli* was used to screen for putative markers indicative of faecal inputs. Initially, amplified products were separated on 5% denaturing polyacrylamide gels. Sufficient resolution to differentiate strains was also obtained using agarose gels when selective AFLP primers contained two or three nucleotide overhangs. In addition, as polyacrylamide gel electrophoresis was laborious for screening large numbers of primer combinations, AFLP-PCR products were subsequently screened on 2% agarose.

AFLP PCR products obtained using selective primers with a single selective nucelotide (+1) were not resolvable on agarose gels and hence it was necessary to revert to polyacrylamide gel electrophoresis. Products amplified using primers containing three selective nucleotides (+3) were clearly resolved and enabled limited strain differentiation, to the extent that a putative marker diagnostic of dairy cattle was identified. Generally, however, fewer bands were obtained with these primers, effectively restricting the probability of locating source specific markers. Primers containing a combination of selective nucleotides were also trialed, for example E-T/M-CTC. The number and resolution of AFLP PCR products from the +1/+3 primer combinations were comparable to those obtained from primers comprising two selective nucleotides. Primers with two selective nucleotides were found to provide optimal banding profiles, defined as follows: optimal profiles were those that exhibited clear resolution of bands as well as sufficient fragments to differentiate isolates. Consequently, the majority of screening was undertaken using primers containing two selective nucleotides.

3.4.3 Locating Markers by AFLP

Significant variation in genomic fingerprint complexity was observed with different primer combinations. *E.g.* when primer E-AG was used in combination with M-AG a highly complex fingerprint was produced, whereas a much simpler fingerprint was generated when the same *Eco* RI primer was used with M-AT. In contrast other *Eco* RI primers (*e.g.* E-AC) consistently produced medium/low to medium/high profile complexities when combined with various *Mse* primers.

3.4.3.1 Putative Genus Specific Marker

The AFLP technique indicated an extensive array of potentially useful genus-specific markers for *E. coli.* Figure 3.2 illustrates a fragment at 640 bp that was common to all the strains analysed using selective primers E-GA/M-CA. It should be noted that *E. coli* was the only genus analysed and therefore it was impossible to ascertain the specificity of the putative fragment.

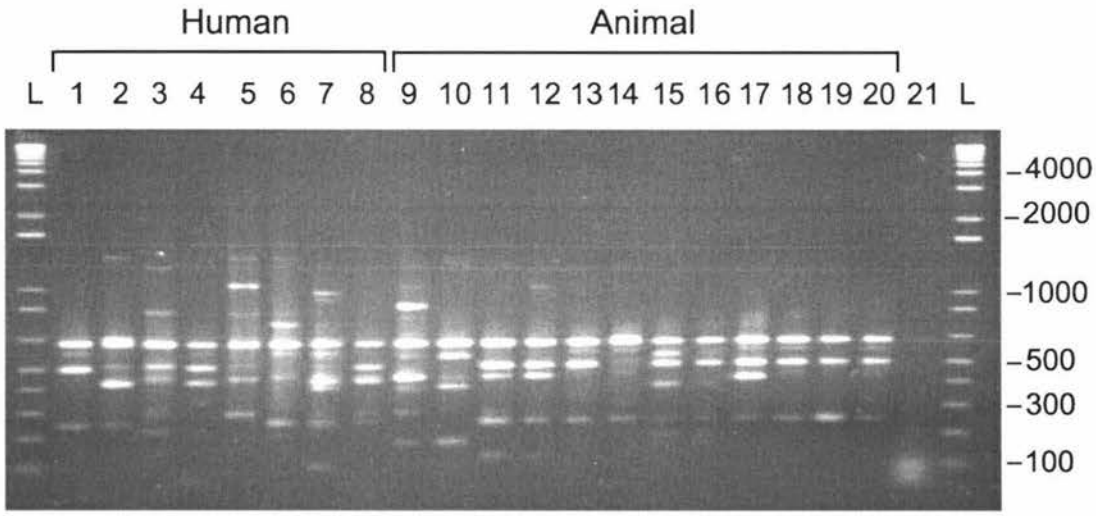


Figure 3.2. Genotypic fingerprints of *E. coli* isolates using AFLP primers E-GA/M-CA. Lane L, 1 Kb Plus DNA Ladder™; Lanes 1 - 4, sewage isolates; Lanes 5 – 8, human stool isolates; Lanes 9 – 12, dairy pond isolates; Lanes 13 – 16, cattle isolates; Lanes 17 - 18, piggery oxidation isolates; Lanes 19 – 20, pig isolates; Lane 21, negative control.

3.4.3.2 Differentiation of Strains

Throughout screening, greater genetic diversity was apparent among isolates from human sources (Figure 3.3; Lanes 1-8) than was observed in either cattle, pig or piggery oxidation pond isolates (Lanes 13-20). Dairy pond isolates (Lanes 9-12) appeared to exhibit intermediate diversity, displaying similarities to the other animal fingerprints but with a greater diversity of fragments. Numerous strains exhibited similar genotypic fingerprints — most often observed amongst animal isolates. For example, the fingerprints generated using primers E-AT/M-AG (Figure 3.3) showed that animal isolates consistently contained a combination of three bands (120, 190 and 250 bp) (Figure 3.3, marked 2-4), which occurred in significantly higher frequencies in animal isolates (92%) than in human isolates (38%). Hence, these bands were not exclusive to the animal strains, as most human strains possessed at least one or two of the three bands, while three strains (38%) contained all three bands. An additional putative genus-specific fragment was also observed at 650 bp (Figure 3.3, marked 1).

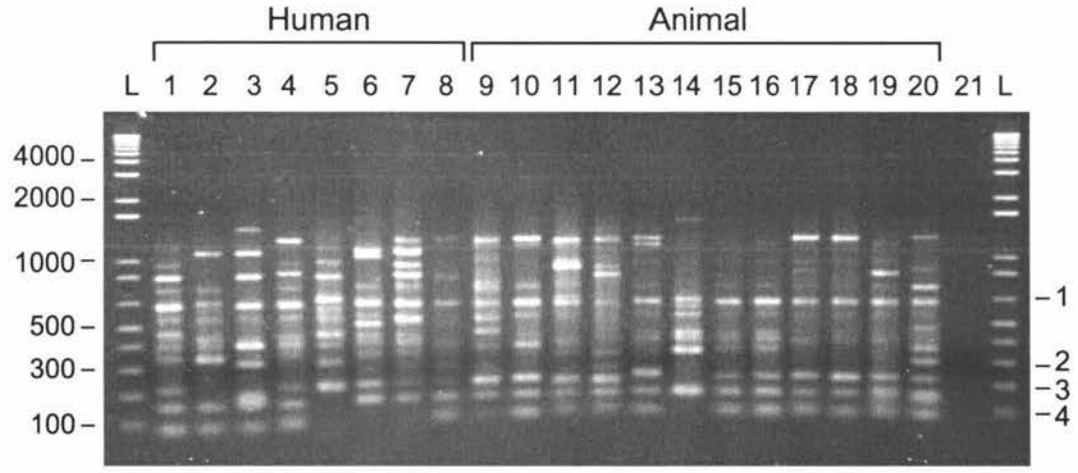


Figure 3.3. Genotypic fingerprints of *E. coli* isolates using AFLP primers E-AT/M-AG. Lane L, 1 Kb Plus DNA Ladder™; Lanes 1 - 4, sewage isolates; Lanes 5 – 8, human stool isolates; Lanes 9 – 12, dairy pond isolates; Lanes 13 – 16, cattle isolates; Lanes 17 - 18, piggery oxidation pond isolates; Lanes 19 – 20, pig isolates; Lane 21, negative control.

A similar observation was noted in profiles generated using AFLP primers E-AGC/M-CT (Figure 3.4). In this instance the triplet band combination was more obvious, and was located at 250, 680 and 780 bp. The diagnostic triplet was present in 63% of animal (non-dairy) strains (Lanes 13, 15, 16, 18, 19). Whereas 38% of human isolates contained only two of the three bands, 250, 680 bp (Lanes 1, 2, 4). An additional isolate (Lane 4) appeared to contain all three bands. The fingerprints obtained from piggery oxidation pond isolates were similar to those from pig stool samples. Interestingly, the profiles from dairy oxidation pond isolates did not represent those from cattle stool isolates.

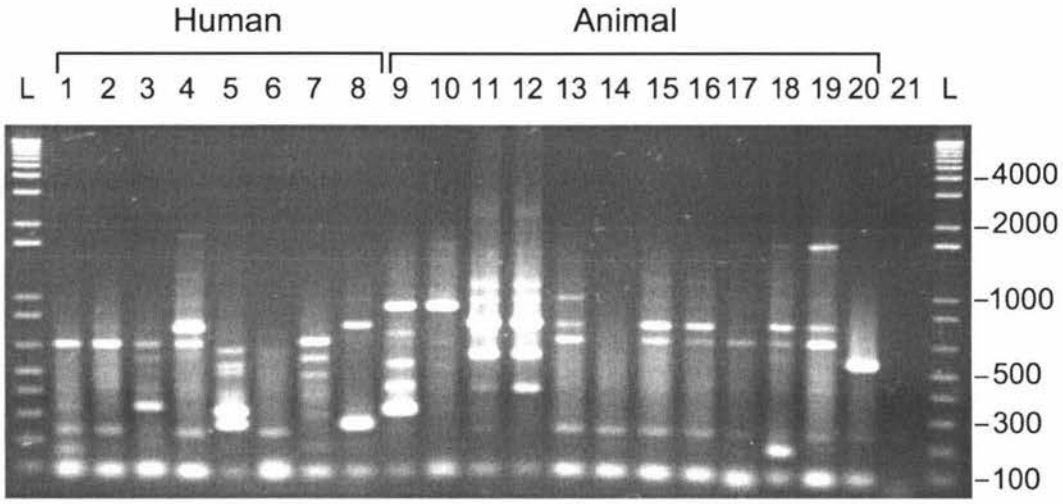


Figure 3.4. Genomic fingerprints of *E. coli* isolates using AFLP primers E-AGC/M-CT. Lane L, 1 Kb Plus DNA Ladder™; Lanes 1 - 4, sewage isolates; Lanes 5 – 8, human stool isolates; Lanes 9 – 12, dairy pond isolates; Lanes 13 – 16, cattle isolates; Lanes 17 - 18, piggery oxidation pond isolates; Lanes 19 – 20, pig isolates; Lane 21, negative control.

3.4.3.3 Putative Marker of Dairy Faecal Strains

The selective AFLP primer combination E-ATT and M-CTA indicated a putative marker diagnostic of isolates from dairy oxidation ponds (Figure 3.5, Lanes 9, 11 and 12). The polymorphic fragment, at approximately 714 bp, was observed in 75% of isolates originating from dairy oxidation ponds. Not surprisingly, it was also observed in a single dairy cattle strain (Lane 13). The fragment however was not

observed in isolates from other sources. A fragment of similar size was observed in a pig manure isolate (Lane 19), however this fragment was 50 bp larger.

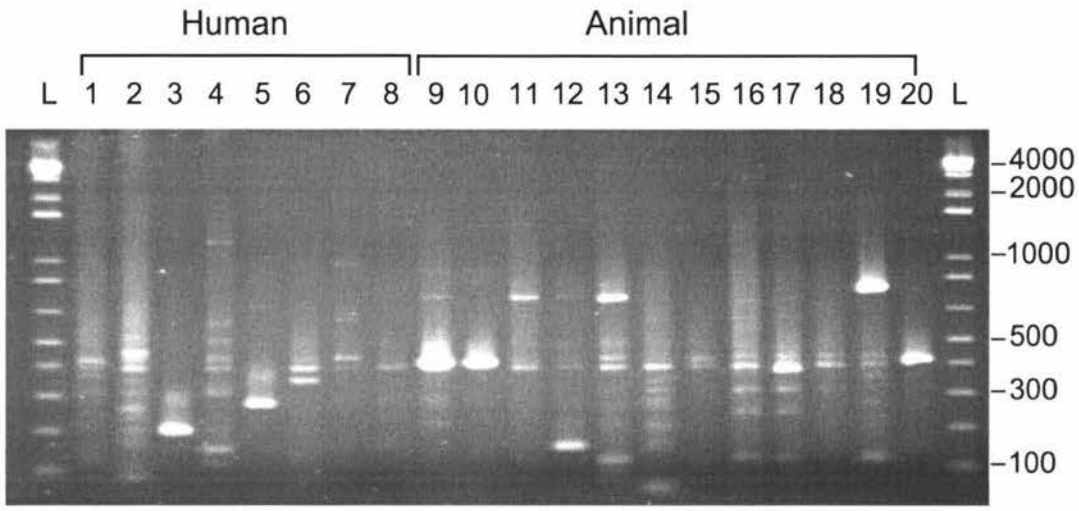


Figure 3.5. Genomic fingerprints of *E. coli* isolates using AFLP primers E-ATT/M-CTA. Lane L, 1 Kb Plus DNA Ladder™; Lanes 1 - 4, sewage isolates; Lanes 5 – 8, human stool isolates; Lanes 9 – 12, dairy pond isolates; Lanes 13 – 14, dairy cattle isolates; Lanes 15 – 16, beef cattle isolates; Lanes 17 - 18, piggery oxidation pond isolates; Lanes 19 – 20, pig isolates.

3.5 DISCUSSION

3.5.1 Resolution of AFLP Products

Polyacrylamide gel electrophoresis using sequencing size gels provides maximum resolution and consequently, enables single nucleotide differences to be resolved. Agarose on the other hand, is rapid, inexpensive and requires less sophisticated equipment. Although fewer fragments are resolved, agarose permits higher sample throughput (Mueller & Wolfenbarger, 1999). Amplified fragment length polymorphisms are typically detected by fluorescence, silver or ethidium bromide staining or, radioactive labelling. Utilisation of fluorescent detection of AFLP products requires a DNA sequencer capable of fluorescent detection. Although this is ideal for automation, DNA sequencers require significant capital expense. Alternatively, detection using radioactive labelling requires containment laboratories, specialised technical expertise and is potentially hazardous. Visualisation using

either silver staining or ethidium bromide is cheap, more easily applied in general laboratories and polymorphic bands may be easily excised for cloning and sequencing. Results obtained in the early stages of research indicated that sufficient resolution was obtained using ethidium bromide staining of agarose gels, such that often all twenty isolates could be uniquely identified depending on the primer combinations used. In addition, several polymorphisms were identified, indicating that differentiation of an isolate's faecal origin might be possible. Subsequent analysis using agarose continued to provide excellent differentiation of strains. Nevertheless, increased resolution would increase the number of fragments obtained and hence increase the probability of locating specific markers.

3.5.2 Application of AFLP to Locating Specific Faecal Markers

The AFLP technique was highly discriminatory — often enabling every strain to be differentiated, even with the reduced resolution of agarose. A number of markers were identified at both the genus and strain level, indicating that the technique may provide a powerful tool for identifying strains from different faecal origins.

3.5.2.1 Putative *E. coli* Diagnosis

Genomic fingerprints generated by AFLP analysis indicated a number of fragments that were common to all the strains analysed. Such fragments have been identified previously in *E. coli* (PE Biosystems, 1996). The markers identified in this research may have potential as genus-specific markers. However, as *E. coli* was the only genus examined it was not possible to ascertain whether the fragments were genus specific or ubiquitous markers common to other organisms. Moreover since several molecular methods are available for discriminating *E. coli* from other enteric bacteria, including the closely related *Shigella* species, it was deemed prudent to focus the research on locating markers that were unique (Sabat *et al.*, 2000).

3.5.2.2 Differentiation of Strains

Strains originating from human stool, sewage and dairy oxidation ponds exhibited greater genetic diversity than isolates from cattle and swine stool samples, or piggery oxidation ponds. Similar results were obtained using the RAPD-PCR primer, GC60. The genetic diversity observed amongst sewage isolates has been reported previously (Pupo & Richardson, 1995). Although the genomic fingerprints from animals often exhibited noticeable similarity, a singular marker indicative of animal strains was not observed using the primer combinations screened here. Instead, doublet or triplet band combinations were commonly observed and were also apparent in the fingerprints generated using RAPD-PCR primers (*e.g.* GC60). The core similarities observed in animal (non-dairy pond) strains, may be useful for assigning isolates to faecal origin. Two approaches could utilise such differences to differentiate the faecal origin of strains. One approach could target doublet or triplet band combinations using target-specific PCR to specifically amplify these fragments. The second approach could utilise the entire AFLP fingerprint in combination with statistical analysis, for assigning strains to faecal origin as described in Chapter Two.

The former approach, where two or three core bands were consistently obtained in the majority of animal profiles should facilitate the development of target-specific PCR primers. Specific PCR amplification of the two or three bands would enable routine monitoring of isolates. Ideally, the primers would be designed for multiplex PCR, in which the diagnostic bands are analysed in a single reaction.

The second approach, using computer-assisted analysis of the entire AFLP fingerprint and statistical analysis, may be used to group isolates according to animal (comprising cattle manure, swine and piggery oxidation ponds) or non-animal faecal origin. Although AFLP analysis is a rapid genotyping technique, in comparison with water quality monitoring protocols the technique is reasonably laborious. Consequently, application of AFLP fingerprinting for routine monitoring is limited for the same reasons as RAPD methods. These include the labour intensive nature of generating fingerprints, increased experimental costs and delayed results (several days to obtain an AFLP fingerprint). In addition, the protocol requires technical

expertise for both analysis and interpretation of results and, for semi-automated analysis, requires expensive DNA sequencing systems. The technique may however be suitable for monitoring "special interest samples" (*e.g.* prosecution cases; blatant non-compliance of consents; sites indicating significant faecal inputs of unknown origin). Even where this approach is used only for "special interest" samples, the AFLP fingerprints must be compared with a reference database comprising genotypic fingerprints of isolates from known sources. Hence the application of this approach must be carefully considered prior to commencing further research, as even the generation of a reference database necessitates significant resource allocation. However, once the database was established and verified, it would provide a valuable resource.

3.5.2.3 Diagnosing Faecal Origin

3.3.2.3.1 Diagnosis of Dairy Strains

The observation of a 715 bp fragment in isolates from dairy oxidation ponds supported the idea that the AFLP technique may provide a suitable method for locating markers diagnostic of faecal origin. The polymorphism was present in isolates of dairy origin, but absent in isolates from beef cattle, other animals and humans. Not surprisingly the marker was also detected in a dairy cattle isolate — the primary source of faecal input to dairy oxidation ponds. Although ideally the marker would be present in all isolates from a particular source, given the variability observed in *E. coli* genotypic fingerprints this seemed unlikely. Moreover, if the polymorphism observed was a function of niche adaptation, it would only be expected to occur in resident *E. coli* isolates. As the marker was present in the majority of dairy isolates and was absent in non-dairy isolates, it was considered a suitable target for further research. Since a single marker was identified, it should be possible to develop a screening method more suited to routine monitoring than the AFLP technique. Chapter Four discusses sequence analysis, development of a method for screening isolates, assessment if it's specificity, and application of the diagnostic marker to environmental monitoring.

3.5.3 Host Specific Association of Strains

The occurrence of polymorphisms specific to particular faecal sources supports the growing evidence that the population structure of *E. coli* may be clustered according to ecological niche (Guttman, 1997; Souza *et al.*, 1999). Turner (1996) reported a polymorphism that correlated with *E. coli* isolates of human origin. Furthermore, antibiotic resistance profiles and REP-PCR fingerprints of *E. coli* also indicate ecological clustering of isolates (Dombek *et al.*, 2000; Harwood *et al.*, 2000; Krumperman, 1983; Lipman *et al.*, 1995; Parveen *et al.*, 1997). Early reports that host association of strains was not apparent (Ochman *et al.*, 1983) may have been a function of insufficient discrimination provided by the techniques utilised.

Host associative adaptations in *E. coli* may arise as a result of selective forces imposed by environmental factors. Such adaptations may be facilitated through niche-specific selection which maintains the adaptive polymorphism (Reeves, 1992). The acquisition of an advantageous allele may expand rapidly through the microbial population, purging genetic diversity at the selected and linked sites — this process is known as "selective sweeping" (Guttman, 1997). If the polymorphism is only advantageous to strains from a particular ecological niche then the selective process will result in the ecological subdivision of the population (Guttman, 1997; Reeves, 1992; Turner *et al.*, 1997a). The polymorphisms reported in this thesis may or may not have constituted the genetic basis for host specific adaptations. Rather, the polymorphisms observed might simply have been "hitchhiker" sequences concurrently transferred to the strain alongside genetic information that conferred host-specific adaptation (Milkman & McKane, 1995; Turner, 1996).

3.5.4 Comparison of Techniques – AFLP versus RAPD Analysis

Both RAPD-PCR and AFLP analysis provided significant differentiation of isolates. Discrimination at the strain level and putative genus discrimination was possible using either technique. The observation that AFLP fingerprints from human isolates were often more diverse than those animal isolates supported the results obtained by RAPD-PCR using primer GC60.

RAPD-PCR was advantageous in its simplicity and rapidity, but low stringency PCR conditions can result in PCR artefacts (Mueller & Wolfenbarger, 1999; Rabouam *et al,* 1999). Although the AFLP technique was slightly more labour intensive than RAPD-PCR, it allowed greater control over experimental parameters. In theory, the number of polymorphisms detected by AFLP analysis is limited only by the power of the separation and the detection system used (Vos *et al.*, 1995). Moreover, the AFLP technique appeared more useful for locating markers indicative of faecal origin, as demonstrated by the identification of a dairy faecal marker and the consistent similarities observed between animal (non-dairy pond) isolates. Similar results have been reported in the literature. Clerc *et al.,* (1998) reported clearer delineation of *Pseudomonas* strains using AFLP rather than RAPD-PCR. Both RAPD-PCR and AFLP analysis would be suited to computer-aided cluster analysis for grouping strains according to faecal origin. However, neither method is conducive to routine environmental monitoring for the reasons mentioned previously, but may be applicable to selected monitoring applications.

3.5.5 Conclusions

The aim of the research described in this chapter was to use AFLP analysis to locate genetic markers indicative of faecal contamination in water. Once a diagnostic marker has been identified, the sequence information should provide a basis upon which to develop screening methodologies (such as target-specific PCR). In summary, AFLP analysis was a useful tool for differentiating strains and for locating markers indicative of the faecal origin of strains. The technique enabled identification of several putative *E. coli* markers and a putative dairy faecal marker. In addition, various combinations of bands were consistently apparent in animal (non-dairy pond) isolates. Although AFLP analysis provided better discrimination of strains than RAPD-PCR, and enabled source-specific markers to be identified, this may have been a function of the number of primer combinations that were trialled with each technique.

Computer-assisted gel analysis of the entire AFLP fingerprint combined with statistical analysis presents an alternative approach for identifying the origin of faecal isolates. Although this protocol is applicable to specific monitoring applications, it is less suitable for routine monitoring than target-specific PCR analysis of the diagnostic marker. In comparison, target-specific PCR can be developed into a relatively fast screening procedure and simply requires presence or absence interpretation of the data, enabling faster and more accurate conclusions.

3.5.6 Future Research

3.5.6.1 Dairy Cattle Faecal Marker

Because the putative dairy cattle faecal marker occurred in higher frequencies than the combination of putative "animal" bands, and because only a single marker appeared to differentiate dairy strains from other sources, the dairy faecal marker was selected as a target for further research. The next step in this research was to:

1. Develop a method conducive to screening isolates for the dairy cattle marker (*e.g.* target-specific PCR).

2. Verify the marker's specificity for isolates of dairy origin.

3. Trial the marker in a field study, to assess the feasibility of using the marker for environmental monitoring.

These experiments are described in Chapter Four.

In addition, the following research could also be investigated.

4. The diagnostic marker may be developed into a colony hybridisation screening procedure. Screening by colony hybridisation would enable a greater proportion of the *E. coli* population to be screened for the dairy cattle faecal marker but would not reduce the analysis time.

3.5.6.2 Alternative Markers

1. The combination of doublet or triplet putative "animal" bands warrants further investigation, hence sequence analysis could be undertaken on these fragments, enabling primers to be designed which facilitate target-specific PCR amplification.

2. Discrimination using the AFLP profile and statistical analysis could be investigated. This technique indicates much promise from the literature and the core similarities observed in this research, suggests that this approach should be feasible. Careful consideration of the medium and long term applicability of this approach prior to embarking on further research, as the resource (time and expenditure) commitment is much higher than developing target-specific markers. Having said that, this approach may enable definition of faecal origin in a single experiment, whereas as application of target-specific PCR primers, may require several screening experiments with different primers identify which faecal sources the isolates originate from.

4 PCR Primer Design for Detection of the Dairy Faecal Marker

4.1 INTRODUCTION

A putative dairy cattle faecal marker was identified by AFLP analysis (Chapter Three). An effective differential indicator must be (1) specific to the faecal source, (2) easily and rapidly detected and (3) present in high frequencies. Although AFLP analysis may be used for monitoring purposes if necessary, it is generally unsuitable for routine screening of large numbers of isolates because the technique is labour intensive and requires complex data interpretation. Instead, a more rapid and more easily interpreted screening protocol, such as target-specific PCR of the diagnostic fragment, would be advantageous. The research described in this chapter developed PCR primers to enable detection of the diagnostic fragment, and also examined the specificity of the marker.

4.1.1 Research Strategy

The research described in this chapter developed a protocol based on target-specific PCR, which enabled isolates to be screened for the putative dairy marker. To determine the specificity and frequency of the marker, research investigated the prevalence of the marker in dairy and non-dairy cattle isolates. In addition, the temporal stability of marker positive strains in a faecal source was also investigated.

4.2 MATERIALS

4.2.1 Chemicals

Unless stated, all chemicals used were analytical grade and were purchased from either BDH Laboratory Supplies (Dorset, UK) or Sigma Chemical Company (St Louis, MO, USA). The following 'Ultra pure' molecular biology chemicals were purchased from Gibco BRL (Life Technologies, Rockville, MD, USA); isopropyl-β-D-thiogalactosidase (ITPG) and 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal). Deionised water, herein described Milli-Q water, was used for preparing all reagents and buffers.

4.2.2 Reagents

4.2.2.1 Buffered phenol, pH 7.9

Phenol saturated with Tris-buffer, pH 7.9, was purchased from BDH Laboratories Supplies.

4.2.2.2 STET Buffer

STET buffer comprised of 10 mM Tris-HCl, pH 8.0, 0.1 M NaCl, 1 mM EDTA, 5% Triton-X 100.

4.2.3 Enzymes, DNA Ladders and Primer Synthesis

Taq polymerase and 10 x *Taq* polymerase buffer (100 mM Tris-HCl, pH 8.3, containing 15 mM $MgCl_2$, 500 mm KCl) were purchased from Roche Diagnostics (Auckland, NZ). Restriction endonucleases and DNA ladders were obtained from Gibco BRL. Primers were synthesised by Life Technologies (Rockville, MD, USA) at 50 nmole scale and standard purity (Table 4.1).

4.2.4 Media

Luria Broth-Millers was purchased from Difco Laboratories (Detroit, MI, USA).

Table 4.1. Sequencing and target-specific PCR primers

Application	Primer Name	Sequence 5' → 3'
Sequencing from pGEM T Easy Vector	pUC/M13 Forward primer	AACCCATTGCGGTCCCA
	pUC/M13 Reverse primer	ACCATGATTACGCCAAG
Dairy Faecal Marker	Dairy 1B	ATTGCACCGCCAGATATTCC
	Dairy 2B	TGCGCCGTTGATTACTCTGC

4.3 METHODS

4.3.1 Media Preparation

4.3.1.1 Sterilisation

Equipment, media and consumables were sterilised as necessary by autoclaving at 121°C for 15 min. Heat sensitive buffers and antibiotics were sterilised by filtering through a sterile 0.2 μm filter.

4.3.1.2 LB Broth

LB broth was prepared by adding 4.5 g NaCl per litre of Luria broth - Millers and was then autoclaved.

4.3.1.3 LB Plates (amp/IPTG/X-gal)

LB plates containing ampicillin/IPTG/X-gal were prepared for the selective growth of transformant *E. coli*. Four and one half grams of NaCl was added per litre of Luria Agar (Section 2.3.1.3) and autoclaved. Ampillicin was added aseptically to the media to a final concentration of 100 μg/mL. Plates were dried to remove excess moisture and then spread with 100 μL 100 mM IPTG and 20 μL 50 mg/mL X-gal (in N,N'-dimethyl-formamide). Plates were freshly prepared.

4.3.2 Sequence Analysis of the Putative Dairy Faecal Marker

4.3.2.1 Location of the Putative Swine Faecal Marker

Isolation of strains, genomic DNA extractions and location of the putative swine marker using RAPD-PCR primer GC60 were performed as described in Chapter Two.

4.3.2.2 Location of the Putative Dairy Faecal Marker

Isolation of strains, genomic DNA extractions and location of the putative dairy marker using AFLP primers E-ATT/M-CTA, were performed as described in Chapter Three.

4.3.2.3 Band Elution from Agarose Gels

Putative faecal-specific markers were located as described in Section 2.3 and 3.3. Since AFLP products were separated on either agarose or polyacrylamide gels, two different band elution protocols were also used. Where sufficient separation was obtained using agarose gels, putative markers were excised and eluted from agarose gels using a QIAquick Gel Extraction Kit (QIAGEN Inc., Valencia, CA, USA) according to the manufacturer's protocol. The concentration of the eluted product was estimated by electrophoresis against Low DNA Mass^{TM} Ladder.

4.3.2.4 Band Elution from Polyacrylamide Gels

The putative faecal source-specific fragment was excised from polyacrylamide gels. To elute the DNA, the fragment was homogenised in 100 µL sterile Milli-Q water and incubated at 4°C overnight. The eluted DNA was reamplifed using the appropriate selective AFLP primers as follows; each PCR reaction contained 2 µL 10 x *Taq* polymerase buffer, 250 µM of each dNTP (dATP, dCTP, dGTP, dTTP) (Roche Diagnostics), 10 pmole of the appropriate selective primer, 1 U *Taq* polymerase and 2.0 µL supernatant from the eluted fragment. The reaction was diluted to a final volume of 20 µL with sterile Milli-Q water. Amplification conditions are outlined in Table 3.4. Average ramp times were 1°C per second in

either a Perkin Elmer GeneAmp® PCR System 2400 or Eppendorf Mastercycler® (Eppendorf, Netheler, Hinz-GmbH, Hamburg, Germany). Re-amplified fragments were analysed on 2% agarose gels. The concentration of the eluted product was estimated by electrophoresis against Low DNA Mass™ Ladder.

4.3.2.5 Transformation of Putative Marker

Target DNA was ligated into the vector "pGEM-T Easy" (Promega Corp., Madison, WI, USA) (Appendix C.1) and subsequently transformed into MAX Efficiency DH5α™ Competent Cells (Gibco BRL) according to the manufacturer's instructions. Transformed cells were identified by blue-white selection on LB plates (amp/IPTG/X-gal). A single overnight white colony was sub-cultured into 5 mL Luria broth containing 100 µg/mL ampicillin and incubated overnight in a shaking incubator at 37°C.

4.3.2.6 Plasmid DNA Extractions

Plasmid DNA was isolated from transformed *E. coli* cells using the following Rapid Boil Lysis protocol (Sambrook & Russell, 2001). One and one half millilitres of an overnight culture was concentrated by centrifugation for 2 min at 15,6000 x g. The supernatant was discarded and the pellet resuspended in 350 µL STET buffer. Twenty-five microlitres of lysozyme (Sigma) solution (10 mg/mL lysozyme in 10 mM Tris-HCl, pH 8.0) was added to the resuspended pellet and then boiled for 40 sec. Cell debris was pelleted by centrifugation at 15,600 x *g* for 10 min at room temperature and was removed with a sterile toothpick. Forty microlitres of 3 M sodium acetate, pH 5.2, and 420 µL iso-propanol were added to the supernatant. The extraction was incubated at room temperature for 5 min, followed by centrifugation for 10 min. The supernatant was removed, the pellet was washed with 70% ethanol, dried at 37°C and re-suspended in 25 µL sterile Milli-Q water.

The insert DNA was verified by *Eco*RI digestion of plasmid DNA. Five microlitres of plasmid DNA were digested with 10 U *Eco* RI (Roche Diagnostics)

in the presence of 1x *Eco* RI buffer (Gibco BRL) for 60 min at 37°C. Insert size was determined by gel electrophoresis as described in Section 4.3.2.7.

4.3.2.7 Agarose Gel Electrophoresis

1.6% agarose gels were prepared with High Strength Analytical Grade Agarose (BioRad Laboratories, Hercules, CA, USA) in 1 x E-buffer (Appendix A). DNA was diluted with loading buffer (Appendix A) and gels were run at 80 V for 60-90 min. A molecular weight ladder, 1 Kb Plus DNA Ladder™, was loaded in adjacent lanes and used to determine product size. DNA was visualised by staining the gel in 100 μg/mL ethidium bromide and photographed under UV light using a 520LP filter on a Fluor-S™ MultiImager (BioRad Laboratories).

4.3.2.8 DNA Purification

Undigested plasmid DNA was further purified for sequence analysis using a phenol/chloroform protocol based on that of Sambrook & Russell (2001). Plasmid DNA extractions were treated with 5 μL of 2 mg/mL RNAse and incubated for 5 min at 37°C. One hundred microlitres of buffered phenol was added to the DNA, mixed and then the aqueous phase removed after centrifugation (15, 600 x*g* at room temperature). This process was repeated, replacing the phenol with chloroform and re-centrifuging. After removing the upper aqueous phase to a clean tube, DNA was precipitated with 0.1 volume of 3 M sodium acetate, pH 5.2, and 2.5 volumes of ice cold ethanol. Samples were incubated on ice for 30 min and then centrifuged at 15,600 x *g* at 4°C for 30 min. The supernatant was discarded and the pellet was washed in 500 μL 70% ethanol and collected by centrifugation for 15 min at 15,600 x *g*. The pellet was resuspended in 10 μL sterile Milli-Q water. DNA concentration was estimated by agarose gel electrophoresis against Low Mass™ DNA ladder.

4.3.2.9 Sequencing Analysis of the Insert

Insert DNA was sequenced in both directions from plasmid DNA using pUC/M13 Forward and Reverse Primers (Table 4.1) according to the following procedure.

Four microlitres of ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready Reaction PreMix (PE Biosystems, Foster City, CA, USA), 1.6 pmole primer and 100-200 ng template DNA were diluted to 10μL with Milli-Q water. The amplification protocol comprised: 1 cycle of 98°C for 5 min; 25 cycles of 98°C for 10 sec, 50°C for 5 sec, and 60°C for 4 min; reactions were held at 4°C. PCR amplification was performed using a Perkin Elmer GeneAmp® PCR System 9700 (PE Biosystems).

Ten microlitres of the PCR product was mixed with 10 μL Milli-Q water, 2 μL 3 M sodium acetate, pH 5.2 and 50 μL ethanol. The reaction was left on ice for 5 min and then centrifuged at 15,600 x *g* for 15 min at 4°C. The supernatant was discarded, the pellet was resuspended in 500 μL 70% ethanol, re-centrifuged for 10 min at 4°C and then the supernatant discarded. Precipitated PCR product was air-dried at 37°C for 1 h. Reactions were electrophoresed and data collected on an Applied Biosystems Model 377™ Automated DNA Sequencer (ABI Prism, PE Biosystems).

4.3.3 Target-specific PCR for Putative Dairy Faecal Marker

4.3.3.1 Primer Design

Sequence alignments for putative source-specific markers were performed using Omiga 2.0 software (Oxford Molecular Ltd, Oxford, England) and primers were constructed with Primer Designer 2.0 (Scientific and Educational Software, Durham, NC, USA). Sequence comparisons were analysed using Blast N 2.1.1, (November 2000) (Altschul *et al.,* 1997). The primers, Dairy 1B and Dairy 2B, designed for target-specific PCR amplification of the putative dairy faecal marker are shown in Table 4.1 and illustrated in Appendix C.2.

4.3.3.2 DNA Lysate Preparation - Rapid Boil Method

DNA lysates were prepared as described previously (de Lamballerie *et al.,* 1992). A single overnight colony was resuspended in 100 μL of 5% Chelex 100 (BioRad

Laboratories), suspensions were boiled for 3 min and then centrifuged at 15,600 x *g* for 5 min. Supernatants were removed and stored at 4°C until required. DNA lysates were used as template DNA to determine the specificity of the primer pair.

4.3.3.3 Target-Specific PCR for Putative Dairy Faecal Marker

Target-specific PCR of the putative dairy faecal marker was optimised for the following parameters: primer annealing temperature, $MgCl_2$ concentration and number of amplification cycles. All optimisation experiments are described in Appendix B.

The optimised conditions for the putative dairy faecal marker were as follows. Reactions contained 2 µL 10 x *Taq* polymerase buffer, 10 pmole each primer (Dairy 1B and Dairy 2B), 250 µM each dNTP (dATP, dCTP, dGTP, dTTP) (Roche Diagnostics), 1.0 U *Taq* polymerase and 1.0 µL DNA lysate. Sterile Milli-Q water was added to a final volume of 20 µL. Positive and negative controls were included in all experiments. The amplification program comprised of 1 cycle of 94°C for 30 sec; 25 cycles of 94°C for 30 sec, 62°C for 30 sec, and 72°C for 30 sec; 1 cycle of 94°C for 30 sec, 62°C for 30 sec, and 72°C for 5 min and held at 4°C until required. PCR amplification was performed using an Eppendorf Mastercycler®. Target-specific PCR products were separated by gel electrophoresis on 1.6% agarose, and were visualised by UV illumination as described in Section 4.3.2.7. 1 Kb Plus DNA Ladder™ was run in adjacent lanes to determine product size.

4.3.3.4 Sequence Verification of Amplified PCR Product

The PCR amplified product was verified by sequence analysis. PCR products from a standard screening reaction were purified using a High Pure PCR Product Purification Kit (Boehringer Mannehim, Roche) according to the manufacturer's instructions. Concentration of the eluted DNA template was estimated using agarose gel electrophoresis against Low Mass™ DNA ladder. Templates were

sequenced in both directions as described in Section 4.3.2.9 using the primers Dairy 1B and Dairy 2B (Table 4.1) and 40 ng DNA template.

4.3.4 Restriction Enzyme Digestion of Amplified Dairy Faecal Marker

A routine method was required to confirm that the amplified PCR products were target-specific. Restriction enzyme digestion of the products was undertaken using the following conditions: Two microlitres of 10U/μL *Pst*I (Amersham Pharmacia Biotech, Auckland, NZ), 2.5 μL 10 x React 2 buffer (Gibco BRL) and 5 μL amplified PCR product were diluted to 25 μL with Milli Q water. Reactions were incubated at 37℃ for 60 min. Digestion products were separated by electrophoresis as described in Section 4.3.2.9.

4.3.5 Marker Specificity of the Putative Dairy Faecal Marker

The aim of this survey was to obtain an initial indication of the marker's specificity. Over two hundred *E. coli* strains, isolated from twenty-one faecal samples, were screened for the dairy faecal marker. Samples were collected from ten different faecal source types and several geographical locations.

Human faecal specimens were obtained from seven healthy human subjects, aged between 20-50 years, who resided the Manawatu and Waikato regions. All subjects were from different households and were apparently genetically unrelated. Samples from aeration lagoons and oxidation ponds were collected near the pond outlets. Freshly excreted faecal specimens were collected from bovine, ovine, swine and equine sources. The geographical origin of each of the samples is described in Table 2.2. All samples were collected in sterile pottles, stored on ice and processed, where possible, within six hours of sampling. Strain isolation is described in Section 2.3.3. Isolates were analysed by target-specific PCR using primer Dairy 1B and Dairy 2B as described in Section 4.3.3.

4.4 RESULTS

4.4.1 Sequence Analysis of the Putative Faecal Markers

4.4.1.1 Fragment Transformation

Band excision from agarose gels (Section 4.3.2.3) proved a more effective method of marker purification than PCR re-amplification of bands excised from polyacrylamide gels (Section 4.3.2.4). Re-amplification from polyacrylamide gels often resulted in amplification of non-specific products and/or low amplification efficiency of target products. Sufficient resolution, to enable band excision, was generally obtained when screening fingerprints from AFLP selective primers containing either two or three selective nucleotide overhangs.

4.4.1.2 Sequence Analysis of the Putative Swine Marker

Quality sequence data was not obtained for the putative swine marker. This was presumed to be a function of the large size of the fragment. As research was continuing on a more promising dairy faecal marker, the swine marker was not pursued further within the time constraints of this research.

4.4.1.3 Identity of the Putative Dairy Faecal Marker

Sequence analysis indicated that the polymorphic fragment was a *Mse*/*Mse* digestion product, 714 base pairs in length (Appendix C.2 illustrates the polymorphic fragment, as well as the AFLP and designed target-specific primers, Dairy 1B and Dairy 2B). Sequence alignment of the polymorphic fragment with the BlastN Sequence Database version 2.1.2 revealed high homology with gene sequences in *E. coli* K12 (93%; E value 0.0) (Appendix C.2). Sequence homology extended from the start of the fragment (base pair position 1) until position 589. The homologous region mapped to the complement of a region in *E. coli* encoding a lambdoid prophage homolog for the tail fibre assembly protein (2521-1938 bp). Sequence homology downstream from this point diverged and consequently the region 590 to 714 bp did not exhibit homology with proximal sequences of *E. coli*. K12 partial homology to the latter region was however

scattered throughout the *E. coli* K12 genome. The reverse primer (Dairy 2B) was located in this region and did not exhibit significant homology with any sequence listed in the genome database. Appendix C.3 illustrates the sequence alignment between the polymorphic fragment and the prophage homolog from *E coli* K12. Sequence information from the polymorphic region was not obtained for marker negative isolates hence it was not possible to identify the specific location of the polymorphism present in dairy and non-dairy isolates. The polymorphism was assumed to be located, in the region that diverged from that of the *E. coli* K12 genome as listed in the BLASTN database.

Homology with Bacteriophage lambda was also present – a fact unsurprising given the homology lambdoid prophage in *E. coli*. The homologous region was the same as that described for the prophage, although with a slightly higher score (96%; E value 0.0). Comparison of the translated sequence, using a BlastP search, mapped to the tail fibre assembly protein and the equivalent homologs in *E. coli*. Sequence homology again diverged downstream of position 589.

4.4.2 Target-specific PCR for Putative Dairy Cattle Faecal Marker

Twenty-mer primers were designed to enable target-specific amplification of the putative dairy faecal marker. The primers, Dairy 1B and Dairy 2B, were designed to amplify a 462 base pair product (Appendix C.2). The reverse primer, Dairy 2B, was located at 600 – 620 bp of the polymorphic region, downstream from where sequence homology diverged with the *E. coli* sequence as described in the BlastN genomic database. Sequence analysis of the PCR product verified amplification of the target region. PCR amplification was optimised for primer annealing temperature, $MgCl_2$ concentration and number of cycles – results are shown in Appendix B. Optimised conditions are described in Section 4.3.3. Figure 4.1 illustrates the products obtained using target-specific PCR and primers Dairy 1B and 2B against a range of faecal isolates. The putative dairy marker was observed at 462 bp in strains originating from dairy faecal sources (Lanes 1-8); however, it was absent in strains from non- dairy faecal sources (Lanes 9-20).

A non-target fragment was intermittently amplified from human stool isolates and rarely amplified from sewage treatment samples (Lane 9). The amplified product was significantly larger (1470 bp) than the target dairy faecal marker (462 bp). Optimisation of PCR amplification failed to eliminate the non-target fragment. One human subject contained a significant number of isolates (80%), which amplified the 1470 bp product. Sequence analysis of the non-target fragment was unsuccessful, presumably due to the large size of the fragment.

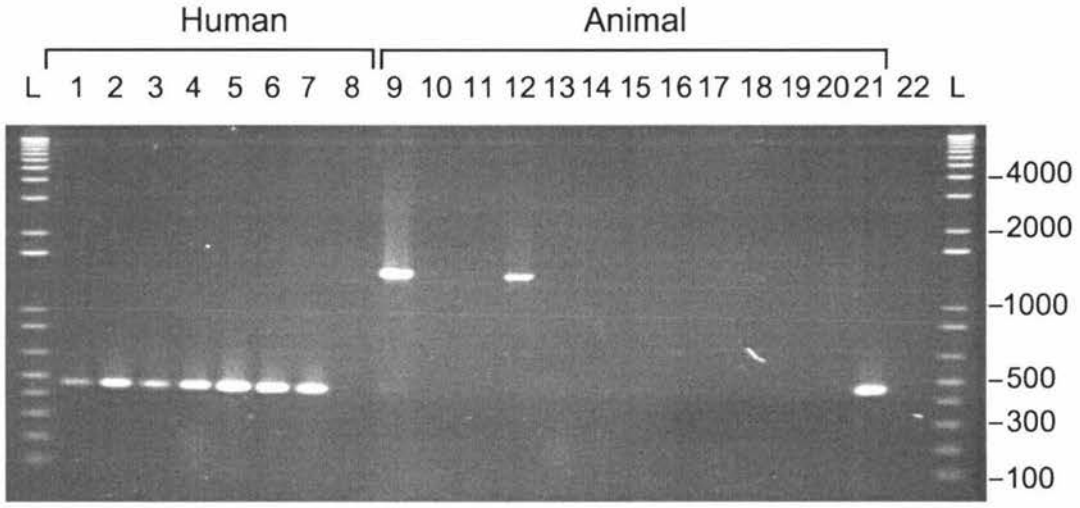


Figure 4.1. Target-specific PCR of the putative dairy faecal marker. Lane L, 1 Kb Plus DNA Ladder™; Lanes 1 - 4, dairy pond isolates; Lanes 5 – 8, cattle isolates; Lanes 9 – 10, sewage isolates; Lanes 11 – 12, human stool isolates; Lanes 13-14, sheep isolates; Lanes 15 – 16 horse isolates, Lanes 17 - 18, pig isolates; Lanes 19 – 20, piggery oxidation pond isolates; Lane 21, positive control - dairy pond isolate D2/4; Lane 22, negative control.

4.4.3 Restriction Enzyme Digestion

Theoretical digestion using Webcutter 2.0 (Max Heiman, copyright 1997) indicated *Pst*I would cleave the PCR amplified fragment at positions 322 and 445 resulting in 3 bands – 17, 123 and 322 base pairs in length (Appendix C.4). Restriction digestion of the amplified PCR product with *Pst*I produced two bands located at approximately 120 and 330 bp (Figure 4.2; Lanes 1-2, & 5). The smaller 17 bp fragment was not observed on agarose gels. Restriction enzyme

digestion of the rarely encountered non-target product gave a similar profile with an additional fragment at approximately 1080 bp (Lane 3 & 4). Dairy pond isolate D2/4 was included as a positive control (Lane 5) as the PCR product had been previously verified by sequence analysis

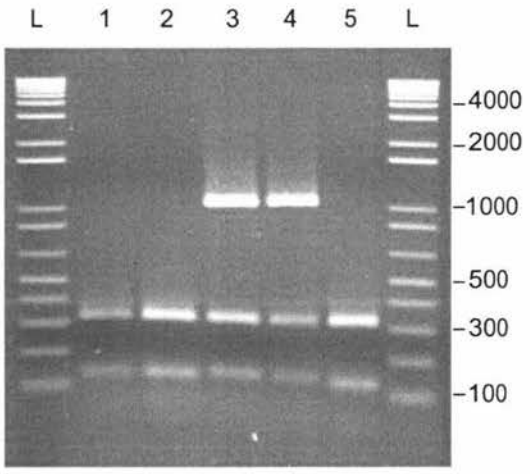


Figure 4.2. *Pst*I digestion of the dairy faecal marker & non-specific product. Lane L, 1 Kb Plus DNA LadderTM; Lane 1, dairy pond isolate D1/1; Lane 2, dairy pond isolate D2/2; Lane 3, sewage isolate HS1/1; Lane 4, human isolate 71/1; Lane 5, positive control - dairy pond isolate D2/4

4.4.4 Specificity of the Putative Dairy Faecal Marker

Marker frequencies from various faecal sources are reported in Tables 4.2 and 4.3. The putative dairy faecal marker was only found in isolates originating from dairy cattle sources, with the exception of one pig faecal isolate that was marker positive. Even isolates originating from beef cattle did not possess the marker. Overall, approximately half of all dairy isolates possessed the marker: 48% of secondary dairy oxidation ponds and 52% of isolates from dairy cattle manure. Of isolates originating from non-dairy sources, 0.9% of isolates were marker positive. Consequently the putative marker was shown to be highly specific. Significant variation was observed in the occurrence of the marker in faecal samples that originated from dairy sources (20-93%). This variation occurred in samples from different sources, geographical locations and over time.

Table 4.2. Specificity of putative dairy faecal marker – dairy faecal isolates

Dairy bovine faecal sources (N[1])	Frequency
Dairy Secondary Oxidation Pond	
Manawatu 1 (10)	50%
Taranaki (14)	93%
Manawatu 2 (15)	20%
Manawatu 3 (pre-spray irrigation) (20)	35%
Average (59)	*48%*
Dairy Primary Oxidation Pond	
Manawatu 2 (12)	33%
Dairy Spray Irrigation Holding Pond	
Waikato (15)	20%
Dairy Cattle Isolates	
Bay of Plenty (10)	90%
Manawatu (15)	27%
Average (25)	*52%*
Overall Average	***43%***

[1] Number of isolates surveyed

4.4.5 Marker Stability Over Time

Marker variation over time was observed in consecutive samplings from the same site. The Taranaki secondary dairy oxidation pond was surveyed approximately two years later in the field study (Chapter Five). Initial sampling from the site suggested high homogeneity of the *Escherichia coli* population. Marker positive isolates dominated the strain composition, with 93% of strains testing positive. Subsequent analysis of the site in January 2001 however, indicated only 30% of isolates possessed the dairy cattle faecal marker.

Table 4.3. Specificity of putative dairy faecal marker – non-dairy faecal isolates

Non dairy bovine faecal sources (N[1])	Frequency
Beef Cattle Isolates - Taranaki (15)	0%
Piggery Primary Oxidation Pond Taranaki (10)	0%
Pig Isolates Bay of Plenty (10)	10%
Sheep isolates Taranaki (15)	0%
Horse Isolates Bay of Plenty (10)	0%
Sewage Secondary Oxidation Pond	
Taranaki (15)	0%
Manawatu (10)	0%
Human Raw	0%
Manawatu – 6 subjects	0%
Hamilton – 1 subject (31)	0%
Overall Average (116)	***0.9%***

[1] Number of isolates surveyed

4.5 DISCUSSION

4.5.1 Location of the Polymorphism in Dairy Faecal Strains

Sequence analysis indicated that the polymorphism diagnostic of dairy faecal strains was a *Mse/Mse* fragment. In many AFLP methodologies, *Mse/Mse* fragments are not visualised, as only the selective *Eco*RI primer is fluorescently or radioactively labelled. Hence only those products incorporating the label are detected (*e.g. Eco*/Mse* or *Eco*/Eco** products[2]). Because *Eco*RI is a rare base cutter *Eco/Eco* fragments would be rare and as a consequence such methods predominantly detect *Eco/Mse* products. The use of silver or ethidium bromide staining however detects DNA fragments regardless of the digestion profile, enabling visualisation of fragments not detected using fluorescent or radioactive methods, such as *Mse/Mse* fragments.

[2] where * represents the labelled primer

The amplification of a *Mse/Mse* fragment is in contrast to a study which reported that although *Mse/Mse* fragments constitute more than 90% of products from *Eco/Mse* digestions, they are not amplified as efficiently as *Eco/Mse* fragments (Vos *et al.,* 1995). The reduced amplification of *Mse/Mse* fragments has been attributed to two factors. Firstly, an inverted repeat at either end may facilitate a stem-loop structure that competes with primer annealing. Secondly, the lower annealing temperature of *Mse*I primers confer less efficient amplification in the presence of *Eco*RI primers. The reason that *Mse/Mse* fragments were obtained in this study was not clear, however several reasons may be responsible. Limited amplification of *Mse/Mse* products has been shown previously (Vos *et al.,* 1995) and as such amplification may occur at low frequency even in the presence of *Eco*RI primers. The polymorphic fragment may have been large enough to prevent or reduce stem-loop formation at the primer-binding site. Vos *et al.* (1995) reported that stem-loop formation would be more difficult in larger fragments. In addition it is not known whether proximal nucleotides may affect stem loop formation. A number of these and other unknown factors may have contributed to the detection of a *Mse/Mse* fragment.

The polymorphism encoding the marker diagnostic of dairy faecal strains was located in the upstream region where sequence homology diverged – *i.e.* at some point upstream from position 1938 bp in *E. coli.* K12 homology with the remainder of the marker mapped to a prophage homolog in *E. coli* K12 encoding a tail fibre assembly protein. Bacteriophage can encode phenotypic factors that may confer a competitive advantage to the host bacterium. *E. coli* O157:H7, for example, contains a bacteriophage encoded toxin (Miao & Miller, 1999). Such virulence factors may enable the host bacterium to extend its host range or enhance its competitive advantage in the environment. In addition, bacteriophage can spread DNA directly to bacterial populations, enabling rapid transfer of elements that may confer a competitive advantage (Miao & Miller, 1999). It is unknown whether the region encoding the polymorphism provided a phenotypic factor that conferred a competitive advantage to the host. *E. coli* is unlikely to express the protein as part of a tail fibre assembly; however, if it did confer a

phenotypic characteristic, for example adhesive attributes, this may aid colonisation to strains. Alternatively the polymorphism may simply be a "hitchhiker" sequence that was concurrently transferred to the strain alongside genetic material that conferred host-specific adaptation (Milkman & McKane, 1995; Turner, 1996). Hence the polymorphism itself may or may not encode host-specific adaptations.

4.5.2 Target-specific PCR

The putative dairy faecal marker was successfully amplified using the target-specific PCR primers, Dairy 1B and Dairy 2B. The presence of the marker in faecal isolates from dairy cattle and its absence in non-dairy isolates indicated that the polymorphism was at least partially "captured" in the region amplified by the designed primers. Target-specific PCR enabled accurate, easy and rapid detection of putative dairy faecal strains.

Restriction digestion of the amplified product provided an effective method for verification of target products. Although restriction enzyme digestion does not guarantee that the product is the intended target, the probability of amplifying a product of the same size and producing the same restriction profile is extremely low. Restriction enzyme digestion of the non-target fragment that was intermittently obtained from human isolates (1470 bp), produced a digestion profile with some similarities to that of the target polymorphism. The similarities indicated that the fragment was related to the target polymorphism, but that one of the primer binding sites was located approximately 1 kilobase further away than was found in the dairy faecal strains. As a result, non-dairy isolates were easily distinguished from dairy strains even on the rare occasion that the non-specific product was amplified.

4.5.3 Specificity of the Dairy Faecal Marker

The putative dairy faecal marker was shown to be specific - it was present in approximately half of the strains isolated from dairy cattle faecal sources and was virtually absent in all non-dairy faecal strains. The apparent specificity of the marker suggests that it may be useful for tracing sources of dairy faecal contamination. One notable observation was the difference in marker occurrence between dairy and beef cattle isolates. In the analyses reported here, the marker was not detected in isolates originating from beef cattle. This may have been a consequence of several factors. In the first instance, beef cattle isolates were taken from a single sample and hence it may be possible that sampling had inopportunely selected a cattle beast that either did not possess the marker, or possessed it in such low frequencies that it was not detected. As a consequence, the results may not be representative of the marker frequency in beef cattle. Therefore further investigation should be undertaken to assess the validity of this result. There are however legitimate reasons to suggest that such stringent host differentiation may be possible, these are discussed further in Section 4.5.3.1. As there was no other indication that the diagnostic marker was non-specific, the single "false-positive" result obtained from a piggery oxidation pond may have be a contaminating strain originating from dairy faecal matter. Further surveys for marker should keep this result in mind to ascertain whether the "false positive" was real.

Ideally, the diagnostic marker would be present in 100% of dairy faecal isolates. However, given the variability of the *E. coli* population present in hosts this is unlikely. The host microbiota consists of both resident and transient strains (Hartl & Dykhuizen, 1984) or, as defined by other authors, majority/minority strains (Hartley et al., 1977). Host specific adaptations are most likely to occur amongst resident and/or majority strains, and therefore if the marker were truly host associated its maximum possible occurrence should be proportional to the frequency of the resident/majority strains.

4.5.3.1 Differentiation of Dairy and Beef Cattle

As mentioned it may be possible that the marker is extremely specific, such that even dairy and beef cattle may be distinguished. Similar results have been reported in the literature (Krumperman, 1983). Multiple antibiotic resistance profiles (MAR), for example, showed that isolates from dairy cattle grouped with other high risk enteric *E. coli* sources (*e.g* humans, swine, and commercial poultry farms). In contrast, grazing cattle clustered with low-risk enteric sources (sheep and wild animals) (Krumperman, 1983).

If additional sampling confirms that the marker is not present in beef strains, then several factors may account for the difference in marker frequency. Whereas the gastrointestinal tract of dairy and beef cattle should not directly differentiate host-specific adaptations, influences such as diet may significantly influence the intestinal environment and hence the bacterial composition. In addition, farm management factors differ significantly between the two farming groups. As such the presence of marker positive isolates in dairy cattle may be a consequence of diet and/or animal to animal transmission.

During winter, dairy herds in New Zealand are commonly fed fermented feed products such as silage or haylage. High counts of *E. coli:*O157 have been reported in poor quality silage (Fenlon & Wilson, 2000). The origin of *E. coli* in silage is presumed to be faecal contamination of the grass prior to cutting, and which has survived limited storage. In contrast Grant *et al.* (1995) reported that enterohemorrhagic *E. coli* were not isolated from grass/legume, or corn/sunflower silage. Increased shedding of *E. coli:*O157 from cattle has been correlated with consumption of corn silage (Heriott *et al.,* 1998). Therefore in light of these results, although silage may present an external reservoir of *E. coli* and facilitate re-inoculation of dairy cattle, it is more likely that survival is insufficient during the six-month fermentation period. Hence the main cause for shedding of *E. coli* from cattle may be that metabolism of silage influences the gastro-intestinal tract in such a manner that marker positive isolates can successfully compete. Dietary

affects have been shown in sheep (Kudva *et al.,* 1997) and cattle (Duncan *et al.,* 2000).

Dairy cattle are typically farmed at higher stocking rates per land area than beef cattle. Higher stocking rates and land irrigation of effluent may provide an external reservoir of faecal strains for re-inoculation. In addition, dairy cattle have regular animal to animal contact during milking. Consequently, dietary factors and animal to animal transmission, either individually or in combination, may facilitate transfer of strains that have adapted phenotypes conferring competitive advantages. As a consequence the strain composition of the dairy herd may become dominated by these strains.

4.5.4 Application to Environmental Monitoring

The high specificity of the diagnostic marker suggests that it may be suitable for environmental monitoring. Marker detection would be applicable for use in water quality laboratories with the addition of minimal equipment, is relatively easy to perform, and results are easily interpreted. In addition, results which were more informative than current water quality methods could be obtained in almost comparable times.

One disadvantage was the variability observed in marker frequency, which has implications for environmental monitoring. To assess the impact of a contaminating faecal source, the frequency of the marker in the original source must be known to enable estimation of the expected marker frequency following dilution in the receiving waters or in the environment. For example, an effluent containing only 20% marker positive isolates would be more difficult to detect in receiving waters than an effluent containing 93% of isolates. Such variability may mean that the marker is less useful in situations where its frequency is low and as such may not be ideal for monitoring all samples. Instead use of the diagnostic marker may be better targeted at specific sites or samples, *e.g.* prosecution cases concerning consistent or excessive consent breaches, or regular

monitoring of dairy effluent/abbattoir discharges. As such the diagnostic marker could still provide a valuable tool for monitoring faecal inputs in the environment.

While the polymorphic fragment exhibited high homology with the bacteriophage lambda genome, marker specificity should not be compromised, as the reverse primer was located in the region in which sequence homology diverged. Nevertheless, researchers should be aware of this homology, particularly if the method were applied to direct PCR amplification from samples. Under such applications, it may be necessary to verify that false positive results were not obtained as a consequence of lambdoid bacteriophage, which is common in the faeces of man (Kameyama *et al.*, 1999) and the environment.

4.5.5 Conclusions

The diagnostic marker for strains originating from dairy cattle appeared specific and was easy to detect. The marker was present in approximately 50% of strains from dairy faecal sources and as such the method shows reasonable potential for use as an environmental monitoring tool.

4.5.6 Future Research

1. To assess the feasibility of using the marker for environmental surveys, field studies should be undertaken in streams and rivers, upstream and downstream of dairy effluent discharges. This would enable assessment of marker detection in environmental samples, test its ability to trace sources of dairy faecal contamination and determine the feasibility of detection following dilution of effluents in streams and rivers. This experiment is described in Chapter Five.

2. Although this survey analysed a significant number of isolates, far greater surveys should be undertaken and should extend the analysis of the faecal types surveyed in this research as well as ovine and avian faecal

specimens, septic tanks and abattoir effluents. The latter may provide interesting results, as the effluent should contain high levels of beef cattle and ovine faecal inputs and low inputs from dairy cattle. A thorough investigation into the marker's presence in beef cattle should also be undertaken. Samples should also be collected from widely distributed geographical locations, to ascertain the variation in marker prevalence from different geographical locations.

3. The specificity of the marker for dairy cattle isolates as opposed to beef cattle isolates should be investigated to clarify whether the distinction observed in this research was valid.

These analyses should be designed to provide statistical validation of the markers specificity and its usefulness in environmental monitoring, as well as provide a more accurate understanding of marker variation between locations, stability over time and implications use of the marker for environmental monitoring.

5 Field Evaluation - Dairy Faecal Marker

5.1 INTRODUCTION

A marker, which was diagnostic of *E. coli* strains isolated from dairy cattle, was utilised in a field study to assess the feasibility of using it as an environmental monitoring tool. Since the diagnostic marker was capable of identifying strains originating from dairy cattle, it may be useful for indication of agricultural run-offs including effluent irrigation, dairyshed wastes and discharges from oxidation ponds. A preliminary field study was undertaken at sites upstream and downstream of a dairy oxidation pond in an attempt to ascertain ease of marker detection, frequency of the marker and its ability to trace sources of faecal contamination in the environment. Prior to the field study however, the protocol was modified to facilitate high throughput screening and the reproducibility of marker detection was ascertained.

5.2 MATERIALS

5.2.1 Chemicals

Unless stated, all chemicals used were analytical grade and were purchased from either BDH Laboratory Supplies (Dorset, UK) or Sigma Chemical Company (St Louis, MO, USA). Deionised water, herein described as Milli-Q, was used for preparing all reagents and buffers.

5.2.2 Enzymes, DNA Ladders and Primer Synthesis

Taq polymerase and 10 x *Taq* polymerase (100 mM Tris-HCl, pH 8.3, 15 mM $MgCl_2$, 500 mm KCl) were purchased from Roche Diagnostics (Auckland, NZ). 1 Kb Plus DNA Ladder™ was obtained from Gibco BRL (Life Technologies, Rockville, MD, USA). Primers were synthesised by Life Technologies at 50 nmole scale and standard purity (Table 5.1).

Table 5.1. Oligonucleotide primers for the dairy cattle faecal marker

Application	Primer Name	Sequence 5' → 3'
Dairy Faecal Marker	Dairy 1B	ATTGCACCGCCAGATATTCC
	Dairy 2B	TGCGCCGTTGATTACTCTGC

5.2.3 Media

The following media were purchased from Difco Laboratories (Detroit, MI, USA); Bacto Peptone, mFc Agar, Nutrient Agar with MUG (4-methyl-umbelliferyl-β-D-glucuronide), Bacto-Agar and Luria Broth-Millers.

5.3 METHODS

5.3.1 Media Preparation

5.3.1.1 Sterilisation

Equipment, media and consumables were sterilised as necessary by autoclaving at 121°C for 15 min.

5.3.1.2 Diluent

A solution of sterile 0.1% Bacto peptone was sterilised and used as a sample diluent.

5.3.1.3 Preparation of Media

mFC Agar; Nutrient Agar with MUG and Luria's Agar – Millers, were prepared as described in Section 2.3.1.

5.3.2 Reproducibility of Marker Detection

5.3.2.1 Replicate Sample Analysis

Replicate sample analysis was performed to assess the variability of marker detection. A single sample was taken from a Manawatu spray irrigation pond (post-secondary oxidation treatment) and analysed in quadruplicate for marker frequency.

Twenty random *E. coli* isolates from each dilution of the quadruplicate series were analysed for the dairy cattle faecal marker.

5.3.2.2 Replicate PCR Detection of a Set of Isolates

To determine the variation in marker detection resulting from DNA template preparation and PCR analysis, a set of twenty *E. coli* isolates were analysed for the dairy cattle faecal marker over five replicate PCR analyses.

5.3.2.3 Isolation of Faecal *Escherichia coli*

Faecal coliforms were enumerated by membrane filtration as described in Section 9222D of Standard Methods for the Examination of Water and Wastewater, 18th Edition (APHA, 1995). Fifty millilitres of sample were mixed by vortex in the presence of 1 g 3mm glass beads (Scientific Supplies Ltd., Auckland, NZ) for 2 min, to ensure that the sample was homogeneously mixed and to eliminate bacterial clumping. Serial quadruplicate dilutions were prepared from the sample using 0.1% peptone broth and filtered though sterile 0.2 μm gridded membranes (Gelman). Membranes were placed onto mFC Agar and incubated at 44.5°C (± 0.2°C) for 24 h, for the identification of faecal coliforms. After 24 h, membranes were transferred to Nutrient Agar with MUG and were incubated for a further 4 h at 35°C. *E. coli* isolates were identified as those colonies exhibiting blue fluorescence under long-wave length ultraviolet light (366nm) (Shadix & Rice, 1991). Isolates were streaked onto Luria Agar – Millers and incubated overnight at 37°C.

5.3.2.4 PCR Amplification - Optimising High Throughput Screening

To facilitate high throughput screening the use of whole cell suspensions as template DNA (Joshi *et al.,* 1991) was investigated. Overnight colonies from Luria Agar-Millers were re-suspended in 100 μL Milli-Q water in a microtitre plate. Target-specific PCR, using primers Dairy 1B and Dairy 2B, was carried out as described in Section 4.3.3.3. Five microlitres of whole cell suspension were added to 15 μL of PCR cocktail. Screening reactions were performed using an Eppendorf

Mastercycler®. The PCR amplification protocol comprised 1 cycle of 94°C for 3 min; 25 cycles of 94°C for 30 sec, 62°C for 30 sec, and 72°C to 30 sec; 1 cycle of 94°C for 30 sec, 62°C for 30 sec, and 72°C for 5 min. This protocol extended the initial denaturation step (94°C) for 3 min. Amplification results using whole cell suspensions were compared with those obtained from DNA lysates (Section 4.3.3.2).

5.3.2.5 Agarose Gel Electrophoresis

PCR products were separated and visualised using gel electrophoresis on a 1.6% agarose gel. Gels were prepared with High Strength Analytical Grade Agarose (BioRad Laboratories, Hercules, CA, USA) in 1 x E-buffer (Appendix A). Five microlitres of each PCR product were diluted with 2 μL loading buffer (Appendix A) and 4 μL sterile water. Gels were run at 80 V for 60-90 min. 1 Kb Plus DNA Ladder™ was run in adjacent lanes for size determination of products. DNA was visualised by staining the gels with 100 μg/mL ethidium bromide and photographed under UV light using a 520LP filter on a Fluor-S™ MultiImager (BioRad Laboratories).

5.3.3 Field Study

Samples for the field study were collected from a Taranaki dairy farm in January 2001. Samples were collected mid-morning from 25 m upstream of the primary oxidation pond (control) and at 50 m and 100 m downstream of the secondary pond outlet. An additional sample was taken from the secondary oxidation pond near the outlet. All samples were collected using sterile bottles, against the water flow and from below the surface. Samples were stored on ice and processed within 6 h. Twenty random *E. coli* isolates from each field sample were analysed for the dairy cattle faecal marker as described in Section 5.3.2.2 using 5 μL of whole cell suspension. Positive and negative controls were included in all experiments.

5.4 RESULTS

5.4.1 PCR Amplification using Whole Cells versus Lysate DNA

Comparison of PCR products generated using whole cell suspensions (Figure 5.1. Lane 1) and Rapid Boil DNA lysates (Lane 2), demonstrated that the whole cell suspension produced equivalent amplification of the target (462 bp) product. Amplification efficiency was enhanced using an initial denaturation step of 3 min compared with a 1.5 min denaturation step, as illustrated by comparing Lanes 1 and 2 with Lanes 6 and 7.

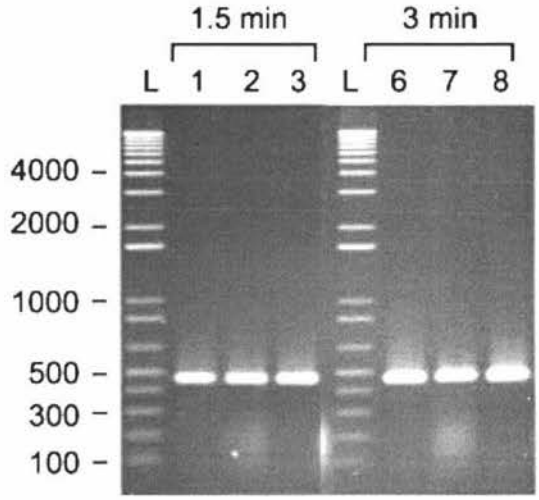


Figure 5.1. Comparison of whole cells versus DNA lysates. Lane L 1 Kb Plus DNA Ladder™; Lanes 1 and 6 were amplified from DNA lysates; Lanes 2 and 7 were amplified from whole cell suspensions; A single strain was used for all reactions. Lanes 3 and 8, positive control.

5.4.2 Reproducibility of Marker Detection

5.4.2.1 Replicate Sample Analysis

Results from the replicate analysis are shown in Table 5.2. Faecal coliform, *E. coli* and marker positive counts were comparable from all four analyses. Detection of the dairy faecal marker was shown to be reproducible, with only minor variation.

Table 5.2. Replicate analysis of marker positive isolates

Replicate	Faecal coliform[1]	*E. coli* [1]	Marker Positive [2]
1	$1.6x10^3$	$1.1x10^3$	35%
2	$1.7x10^3$	$1.4x10^3$	45%
3	$1.4x10^3$	$1.4x10^3$	30%
4	$1.5x10^3$	$1.4x10^3$	35%
Avg	*$1.55x10^3$*	*$1.33x10^3$*	*36%*

[1] (cfu/100 mL); [2] Isolates positive for the dairy marker as a percentage of *E. coli*

5.4.2.2. Replicate PCR Detection from a Set of Isolates

To determine if the minor variability observed in the replicate sample analysis resulted from template preparation and/or PCR amplification, quintuplicate analysis of a set of twenty *E. coli* isolates was undertaken. Results from the replicate analysis were virtually identical from all analyses. Marker positive results were as follows (30%, 30%, 30%, 35%, 30%).

5.4.3 Field Study

Faecal coliform, *E. coli* and marker positive frequencies from each of the sampling sites (Section 5.3.3), are reported in Table 5.3. Faecal coliform and *E. coli* counts from the secondary oxidation pond were comparable with those obtained previously at the site, but were somewhat higher. Enumerations of faecal coliforms and *E. coli* were not elevated downstream of the oxidation ponds as anticipated. In fact of the stream samples the highest bacterial loading was obtained from the upstream site. One of the major factors presumably contributing to this result was the lack of effluent discharge from the pond at the time of sampling. This issue is discussed further in Section 5.5.3.2.

The dairy cattle faecal marker was easily detected in both the effluent and stream samples, with frequencies ranging between 15 and 30%. Figure 5.2 illustrates marker frequency at each of the four sites. Marker frequencies paralleled bacterial

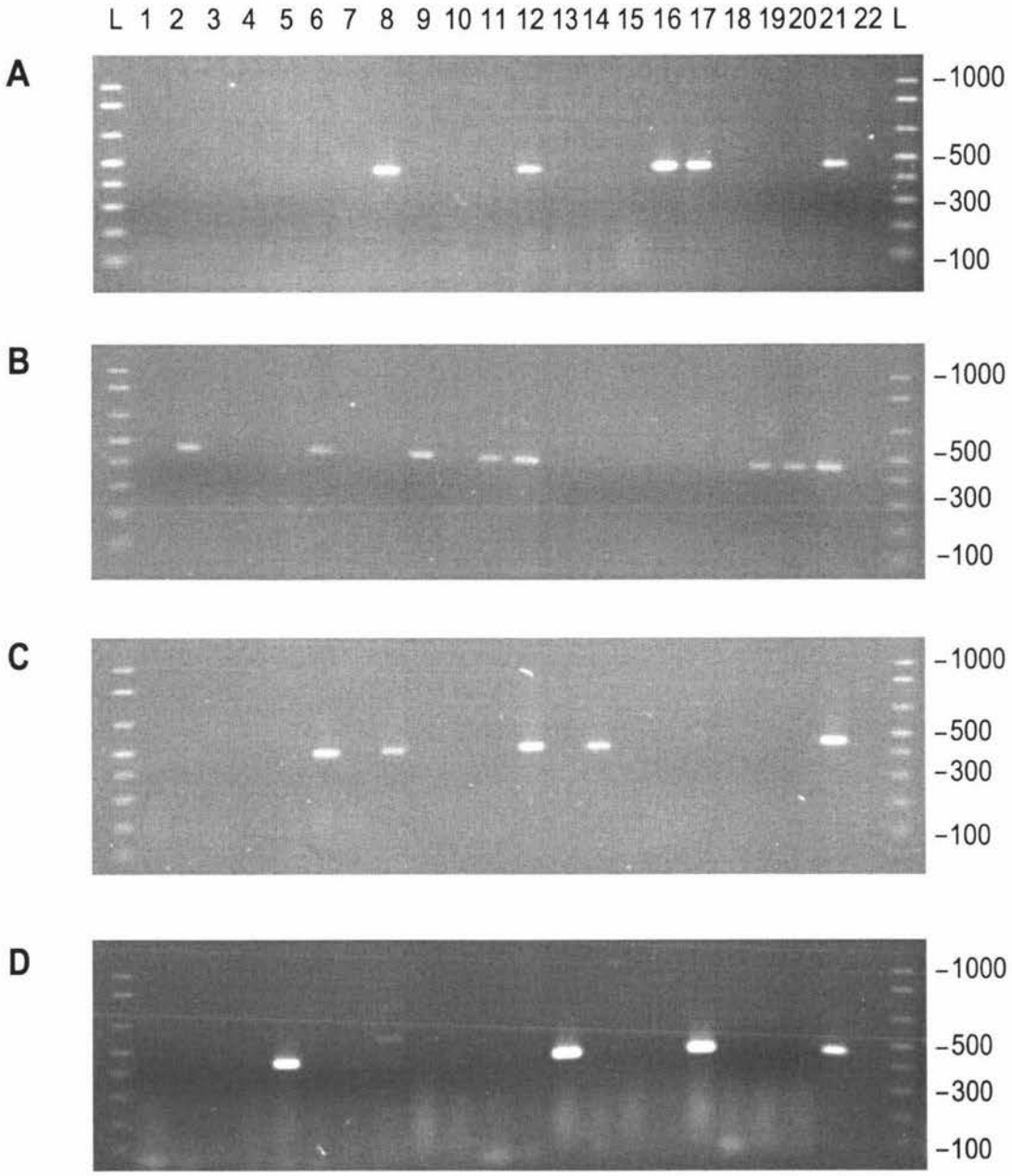


Figure 5.2. Dairy cattle faecal marker frequency at each of the field sites. Lane L, 1 Kb Plus DNA Ladder™; **(A)** Upstream site; **(B)** Dairy Pond Outlet; **(C)** 50m downstream; **(D)** 100m downstream. Lanes 1-20, *E .coli* isolates; Lane 21, positive control; Lane 22, negative control.

enumerations, with the dairy pond effluent exhibiting the highest marker incidence. Frequencies of the dairy marker obtained at the sites upstream and immediately downstream of the ponds were similar, and did not indicate elevated levels at the downstream site as anticipated. This presumably related to the lack of effluent discharge at the time of sampling. Marker frequency declined further downstream, as would be anticipated by dilution and die-off effects.

Table 5.3. Enumeration of faecal coliforms, *E. coli* and marker positive isolates at each of the field sites.

Sampling Site	Faecal coliform[1]	*E. coli*[1]	% *E. coli*[2]	Marker Positive
Upstream	8×10^2	5.7×10^2	71.3%	20%
Outlet	5.8×10^3	4.4×10^3	75.9%	30%
50 m Downstream	6.0×10^2	3.7×10^2	61.2%	20%
100 m Downstream	6.9×10^2	2.5×10^2	36.2%	15%

[1] cfu/100 mL; [2] *E. coli* as percentage of faecal coliform count.

5.5 DISCUSSION

5.5.1 Whole Cell Suspensions versus DNA Lysates

The use of whole cell suspensions as the template for target-specific PCR facilitated rapid screening of large numbers of isolates. The method was shown to be reproducible and indicated no loss in amplification efficiency compared with the use of DNA lysates. These results were in agreement with the findings of Joshi *et al.,* (1991) who reported equally efficient amplification of the PCR product from both *Salmonella typhimurium* and *E. coli* using either purified chromosomal DNA or whole cells from plates or liquid cultures. The increase in initial denaturation time would presumably aid cell lysis, thereby improving consistency in the release of DNA to the PCR cocktail from resistant strains. Resuspension of isolates in sterile deionised water prior to addition to the PCR cocktail minimised agar transfer that has been shown to inhibit PCR amplification (Gibb & Wong, 1998).

5.5.2 Replicate Analysis of Sample

Detection of the dairy cattle faecal marker was reproducible. Minor variation was observed in the replicate sample analysis (Section 5.4.2.1), but this was not due to either template preparation or PCR amplification, as shown by the reproducibility of marker detection during replicate analysis of a set of isolates (Section 5.4.2.2). Variability during replicate sample analysis was probably a consequence of the sample size analysed for the marker. Utilising PCR amplification for detection of the dairy polymorphism immediately constrains the size of the *E. coli* population that may be analysed. Small sample sizes would exacerbate variability in the frequencies observed. Consequently, given the small sample size screened for the marker, the variation in marker incidence was relatively small.

The numbers of isolates screened here were comparable with those reported previously for a marker diagnostic of sewage contamination (Turner, 1996). However the incidence of the dairy cattle faecal strain polymorphism was approximately double that of the sewage-specific marker in environmental samples. Turner (1996) utilised binomial estimations to predict the required sampling size for detecting the sewage marker. The probability of detecting at least one marker positive isolate was 92% when 24 isolates were analysed, and 99% if 45 isolates were analysed. These estimations were based on a marker incidence of 10%. Given the implications for sample handling, processing and cost, the difference between detecting at least one marker positive isolate 92% and 99% of the time is not significantly advantageous. Turner's binomial estimation essentially provides sampling sizes suitable for determining presence/absence indication of the sewage marker. However, quantitative analysis of faecal markers and hence inputs, would be advantageous. Hence further studies would be required to ascertain the optimum isolate sample size per environmental sample, to maximise detection sensitivity and to minimise variability. The financial and logistical constraints of screening large numbers of isolates must also be considered when planning environmental surveys.

5.5.3 Field Study

5.5.3.1 Feasibility of the Detection System

The field study indicated that the marker detection system was a feasible environmental monitoring tool. The analysis was relatively rapid, marker positive isolates were clearly distinguished from marker negative isolates, and frequencies were sufficiently high to enable easy detection. Diagnostic results were obtained within 48 h. This could however be reduced to 32 h, which is almost comparable with standard methods for the enumeration of faecal coliforms and *E. coli,* which take minimum of 24 to 28 hours respectively. The reduced analysis time should be achievable by eliminating isolate purification on Luria agar. Detection of false positives should not be affected, as the dairy cattle strain polymorphism did not indicate sequence homology with any non-target strains that grow on mFc agar.

Further reduction in analysis time could be obtained by eliminating isolate cultivation altogether. The advent of quantitative real-time PCR (Isono, 1997) means that the marker could be amplified directly from environmental samples to give quantitative results. However direct amplification typically requires extensive sample cleanup procedures since PCR inhibitors, such as humic acids, are common in environmental samples (Abbaszadegan *et al.,* 1993; Kreader, 1996; Lewis *et al.,* 2000; Tsai *et al.,* 1993). In addition, non-viable isolates would also be detected and may skew disease risk analyses. Finally, significant capital outlay would be required to purchase the necessary equipment. The latter issue may be addressed by contracting the analysis to laboratories with quantitative real-time PCR capabilities.

5.5.3.2 Diagnosis of Faecal Source

The highest counts of faecal coliforms, *E. coli* and marker positive isolates were recorded from the dairy pond effluent. Enumeration of faecal coliforms and *E. coli* did not indicate elevated faecal loading downstream of the oxidation ponds. Nevertheless, marker positive isolates were obtained in reasonable frequencies at the downstream sites. The dairy marker frequencies were higher than those

reported previously for a diagnostic marker of sewage contamination (Turner, 1996). In the Wellsford study, Auckland (Turner, 1996), the marker (HH1) was identified by target-specific PCR in 8.3% of isolates from both the final sewage effluent and 3 m downstream of the discharge, however the marker was not detectable 50 m downstream. In comparison, the dairy cattle marker was detected in 30% of effluent isolates and 20% of isolates from the site 50m downstream.

5.5.3.2.1 Specificity of Dairy Faecal Marker

In this preliminary field study, the origin of the marker could not be clearly attributed to the dairy effluent as results indicated an additional, but significant, upstream source of marker positive isolates. Several issues must therefore be addressed. In the first instance it may be possible that the marker was not specific to strains originating from dairy faecal sources, as there was no increase in marker incidence downstream from the dairy pond. In support of the marker's validity, previous results indicated that the marker was highly specific (Section 4.4.4). One faecal source, however, that has not been investigated to date is avian. Isolates originating from birds can contribute significant faecal loading to waterways (Alderisio & de Luca, 1999; Oshiro & Fujioka, 1995; Roll & Fujioka, 1997). However given the specificity of the marker, that dairy and beef cattle may be distinguishable, homology with avian isolates is less probable. In addition, significant numbers of birds would be required to affect the stream in a manner comparable with a dairy oxidation pond. Flocks of birds are not typically observed on small streams, similar to the one surveyed, and there were no lakes, dams or ponds upstream of the sampling sites.

5.5.3.3.2 Putative Upstream Source of Dairy Faecal Isolates

While there were no apparent point sources of faecal contamination within the vicinity of the upstream site, several non-point sources may have accounted for the elevated upstream counts. In the first instance, the dairyshed was on a hill immediately above the upstream sampling site. Depending on the efficiency of the wash water collection system, this may have contributed faecal loading to the stream. Secondly, the adjacent farm utilised a powerful spray irrigation system that

may have resulted in faecal loading of the stream, either from direct input (the spray had been shown to travel beyond the stream) or indirect input. Indirect input may have occurred through seepage or excessive irrigation causing pooling and surface runoff. It is not known whether the system had been operating within the vicinity of the stream in the days preceding sampling.

Additionally, while no dairy cattle were apparent in the paddocks immediately upstream, it is possible that these had been present prior to sampling and that the bacterial loadings recorded were the result of delayed filtration through the soil to the stream (Howell *et al.,* 1995; Rothmaier *et al.,* 1997). Reports have indicated significant effects on stream quality as the result of stock access to streams, and fencing has been shown to significantly reduce faecal coliform loading (Hagedorn *et al.,* 1999; Horizons Manawatu-Wanganui, 1999; Howell *et al.,* 1995).

5.5.3.3.3 Absence of Discharge

The lack of discharge at the time of sampling, was presumably one of the most significant factors influencing the inability to conclusively attribute the origin of the dairy faecal marker to the secondary oxidation pond. The lack of discharge was appraised at the time and it was considered that while the ponds were unlikely to continuously discharge, recent effluent discharges should be detected. In addition, monitoring agencies typically would not restrict sampling regimes to coincide with discharge events (Nix *et al.,* 1994). Hence it was determined by the researcher, that sampling in the absence of a concurrent discharge may represent typical environmental surveys. Additionally, in the event that inconclusive results were obtained, the field study could be repeated to coincide with the discharge, thereby providing data pre- and post-effluent discharge.

While sampling may have represented typical environmental surveys, the inconclusive results that were obtained, limited interpretation of the results. Recent dairy contamination was detected, but it was not possible to conclusively ascertain the origin. Had the marker been absent from the upstream site, or at least present in lower frequencies, the specificity of the marker for dairy faecal sources should have

been more easily defined. Instead the high upstream counts, exacerbated by the lack of discharge, may have masked the impact of oxidation ponds on the downstream environment.

Additionally, the site immediately downstream of the discharge point may have been located too far downstream to enable detection of the "peak" in marker incidence ensuing from recent discharges. In the Wellsford study the sewage marker (HH1) was no longer detectable 50 m below the discharge point (Turner, 1996). However, in this field survey the dairy faecal marker was detectable at least 100 m downstream. The peak in marker frequency would be expected to move progressively downstream over time once the discharge ceased. The rate of downstream movement and eventual disappearance of faecal isolates would ultimately depend on the effluent loading, as well as dilution effects and stream flow. The latter factors may also have influenced the results, depending on the time interval and flow since the previous discharge event.

5.5.3.3 Resolving the Inconclusive Results

Due to the confounding issues of the field study, a second survey was planned to clarify the marker's ability to trace cattle faecal contamination in the environment. Preparations to repeat the field study were undertaken, however upon approaching several farmers in the Manawatu region, all reported that they no longer discharged treated effluent to waterways and instead utilised land-based irrigation for effluent disposal. In 1999, 60% of farmers in the Manawatu region utilised land-based effluent disposal and this declining trend in effluent disposal to waterways was anticipated to increase (Horizons Manawatu-Wanganui, 1999). Consequently further field studies may be better aimed at tracing marker positive isolates from land irrigation of effluent. The combination of reduced incidence of discharge to waterways and time constraints meant that the field study was not repeated.

An alternative approach could have utilised "ideal site selection". In such surveys, sites are selected where upstream environments are pristine with no agricultural

inputs, such as in upper catchment areas. This latter option was avoided in the first instance as it applies the marker in a situation rarely encountered in environmental surveys. However for the purposes of an initial survey such approaches may be beneficial.

5.5.4 Conclusions

Detection of the dairy cattle faecal marker was reproducible, relatively rapid and easily applied to environmental monitoring. The marker shows significant promise as a source specific indicator, however further field studies should be undertaken to verify its specificity in the environment.

5.5.5 Future Research

Given the feasibility of the dairy cattle faecal marker as an environmental monitoring tool and its apparent specificity to date, further investigations are recommended.

1. Since analysis of the dairy marker is not conducive to the screening extensive numbers of *E. coli*, the optimum size of the *E. coli* sample population to be monitored should be determined. This would enable monitoring agencies to accurately estimate marker frequency in the environment while minimising labour and expenditure resources.
2. Field studies should be undertaken to assess the ability of the dairy marker to trace dairy cattle faecal contamination in the environment. Such studies could be undertaken following land application of dairy effluent. Both adjacent streams and groundwater should be monitored for the dairy faecal marker, pre- and post-effluent irrigation. As such, several groundwater bores would be required at various locations in the pasture. Moreover, knowledge of the groundwater hydrology at the sampling sites would also be necessary.

6 General Conclusions

The aim of the research described in this thesis was to locate a marker(s) indicative of the origin of faecal contamination in waterways, particularly agricultural inputs. The ability to define the nature of faecal contamination would enable monitoring agencies to trace the cause of the pollution more accurately and hence increase the opportunity to remediate the problem. Chapters Two through to Chapter Five present and discuss the results from this research. This final chapter provides a summary of the findings and their implications and recommendations are made for further research.

6.1 MOLECULAR PROFILING FOR MARKER LOCATION

RAPD-PCR and AFLP analysis provided excellent discrimination of strains. Both techniques were rapid, cost effective and relatively easy to perform. Although AFLP analysis was more laborious than RAPD methods, a greater number of putative diagnostic markers were located. Putative genus-specific markers were identified using both techniques. However, as various molecular methods are currently available for identifying *E. coli* (Fricker & Fricker, 1994; Green *et al.,* 1991; Sabat *et al.,* 2000) these markers were not investigated further. A marker diagnostic of isolates originating from dairy cattle faecal sources was identified and a method suitable for environmental screening of isolates was developed. The identification of such a marker supports the growing evidence that the *E. coli* population structure may encompass extensive ecological clustering (Guttman, 1997; Souza *et al.,* 1999). Turner (1996) identified a marker that correlated with clinical isolates originating from sewage, whereas Guttman (1997) found that strains were more strongly associated by host and origin than would be expected by chance. In addition, REP-PCR analysis indicated strong host association of faecal strains (Dombek *et al.,* 2000), however the technique was not suitable for routine environmental monitoring. Consequently the elucidation of markers diagnostic of sewage (Turner, 1996) and of

dairy cattle (this study) suggests that it would be feasible to construct a library of markers diagnostic of faecal origin. This would enable differentiation of a wide range of contaminating sources, which might include faecal types such as piggeries, septic tanks, avian or ovine sources.

6.1.1 Differentiating Sources of Faecal Contamination

6.1.1.1 Dairy Cattle Faecal Marker

A putative dairy cattle faecal marker was identified by AFLP analysis. PCR primers were designed to amplify the fragment. The dairy cattle faecal marker was highly specific. The diagnostic marker was, however, present only in approximately half of the dairy isolates. Although lower than hoped, this still provided adequate detection of dairy faecal matter. Target-specific PCR enabled rapid, reproducible and conclusive detection of dairy faecal isolates. In addition, the results were easily interpreted. This diagnostic marker is suitable for routine environmental monitoring. However, application may be restricted to "special interest sites" (*e.g.,* consistent and significant water quality deterioration at a particular site for unknown reasons; significant pollution incidents suspected to be caused by faecal contamination; blatant non-compliance of resource consents; prosecution cases).

One of the advantages of the detection protocol in this research is that it can be appended to standard agar-based methods for testing the microbiological quality of fresh waters (*e.g.,* mTEC plus the urease test; mFc followed by nutrient agar with MUG). For example, after complaint samples are analysed for *E. coli* or faecal coliforms by membrane filtration methods, a decision, based on the enumerative results, may be made as to whether it is necessary to further characterise the origin of faecal contamination. Definition of faecal origin could be obtained in an additional eight hours of laboratory time.

6.1.1.2 Animal Markers

Combinations of several bands were consistently apparent in animal (non-dairy pond) isolates in both the AFLP and RAPD-PCR fingerprints. Further research could develop target-specific PCR primers to amplify these bands.

6.1.1.3 Other Markers

A number of other putative markers were apparent. These were not pursued further because isolates from that source were obtained from a single sample (*e.g.* swine isolates). Consequently, a library of markers could be developed to diagnose a range of faecal sources. AFLP analysis was shown to provide a suitable protocol for detecting such markers. RAPD-PCR methods may also be suitable; the lack of diagnostic markers observed with this technique was assumed a function of the limited number of primer combinations that were screened.

6.1.1.4 Differentiation Based on Genotypic Fingerprints

Both RAPD-PCR and AFLP analysis showed that human isolates encompassed greater genetic diversity than animal (non-dairy pond) isolates. Hence animal and human isolates could be generically delineated into two groups because of the restricted genotypes observed in animal isolates. Combining this observation with statistical analysis presents an alternative approach to identifying the origin of faecal isolates. Differentiation of individual sources (*e.g.* dairy oxidation ponds versus piggery oxidation ponds) may be possible with complex fingerprints. These approaches would provide a robust classification system, as isolates are assigned to each group on the basis of several polymorphisms rather than just one.

Although this protocol is applicable to specific monitoring applications, it is less suited than target-specific PCR of diagnostic fragments for the following reasons. RAPD-PCR and AFLP analysis are more time consuming as high quality genomic DNA is required, and numerous manipulations are required for AFLP analysis. Both AFLP and RAPD-PCR analysis require interpretation of complex data, statistical analysis and comparison with a reference database of genomic fingerprints. The

development of a reference database may be laborious for limited application of the technique.

6.2 FUTURE RESEARCH

6.2.1 Recommendations for Monitoring Prevalence of the Dairy Cattle Faecal Marker

1 The specificity of the marker diagnostic for *E. coli* from dairy cattle faecal matter should be verified using a larger number of isolates from a range of faecal sources and geographical locations. This survey should extend the investigation undertaken to date (including ovine, bovine, equine, and swine sources), as well as other sample types such as septic tanks, abbatoirs, *etc*...

2 The field study described in this research was unable to definitively attribute the origin of marker positive isolates to the dairy oxidation pond hence a suitably designed field study is still required. The field study would clarify the utility of the method for tracing sources of dairy faecal contamination in the environment. The field trial may replicate the experiment described here, but should include a site that is discharging at the time of sampling and preferably where there is minimal upstream dairy inputs. Alternatively, a groundwater survey may be undertaken pre- and post-effluent irrigation of dairy shed waste onto pasture.

3 The application of target-specific PCR for environmental monitoring would require an assessment of optimum sampling size (*i.e.,* number of isolates to screen per sample) to accurately estimate the frequency of the marker, while minimising analysis time and cost.

4 An alternative detection method using colony hybridisation and chemiluminescent detection could be developed to enable the dairy marker to be more easily quantified. Although this may not reduce analysis time, and

hazardous chemicals are required, greater numbers of isolates can be screened for the dairy faecal marker. The utility of this method would depend on how unique the diagnostic fragment was, as the probe may hybridise to fragments that are similar but not necessarily identical. Target-specific PCR in contrast is more sensitive to minor sequence differences.

5 In addition, it would be valuable to investigate combining detection of the diagnostic dairy faecal marker with the fragment which is diagnostic for sewage isolates (Turner, 1996). This may be achieved by using methods such as multiplex PCR, which may be used to detect both diagnostic markers in a single reaction. Some difficulties do arise however, including optimising annealing temperature of different primer sets.

6 Finally, the putative swine marker identified by RAPD-PCR could be re-investigated. However, as AFLP analysis indicated a number of other more specific putative faecal markers (including swine), research would be better targeted at these markers.

6.2.2 Recommendations for Further Markers

The protocol outlined in this research is suitable for locating additional diagnostic markers, such as piggery effluents, ovine faecal matter, or even dairy cattle. A second dairy cattle marker used in combination with the one described in this research may enhance the frequency with which dairy isolates can be detected.

6.2.3 Discrimination of Faecal Source Using Genotypic Profiles

Profiling on the basis of genotypic fingerprinting and statistical analysis could be investigated. However, careful consideration of the medium to long term applicability of this approach should be assessed first, as reference databases must be established and ongoing labour intensive protocols are required.

References

Aarts H.J.M., van Lith L.A.J.T. & Keijer J. (1998). High resolution genotyping of *Salmonella* strains by AFLP-fingerprinting. *Letters in Applied Microbiology.* **26**: 131-135.

Abbaszadegan, M., Huber, M., Gerba, C. & Pepper, I. (1993). Detection of enteroviruses in groundwater with the Polymerase Chain Reaction. *Applied Environmental Microbiology.* **59**: 1318-1324.

Alderisio, K.A. & DeLuca, N. (1999). Seasonal enumeration of fecal coliform bacteria from the feces of ring-billed gulls (*Larus delawarensis*) and Canada geese (*Branta canadensis*). *Applied Environmental Microbiology.* **65**: 5628-5630.

Allsop K. & Stickler, D.J. (1985). An assessment of *Bacteriodes fragilis* group organisms as indicators of human faecal pollution. *The Journal of Applied Bacteriology.* **58**: 95-99.

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

APHA. (1995). *Standard Methods for the Examination of Water and Wastewater.* (Franson, M.A.H., Greenberg, A.E., Clesceri, L.S. & Eaton, A.D., Eds.), 19th Edn. APHA, AWWA & WEF, Washington D.C.

Angulo, F. (1999). Adverse human health consequences of antimicrobial use in food animals. *The Infectious Disease Review for the Medical, Veterinary and Environmental Professions.* **1**: 217.

Armon, R. & Kott, Y. (1995). Distribution comparison between coliphages and phages of anaerobic bacteria (*Bacteroides fragilis*) in water sources and their reliability as fecal pollution indicators in drinking water. *Water Science and Technology*. **31**: 215-222.

Arnold, C., Metherell, L., Clewley, J.P. & Stanley, J. (1999). Predictive modelling of fluorescent AFLP: a new approach to the molecular epidemiology of *E. coli*. *Research in Microbiology*. **150**: 33-44.

Arnold, C., Metherell, L., Willshaw, G., Maggs, A. & Stanley, J. (1999). Predictive fluorescent Amplified-Fragment Length Polymorphism analysis of *Escherichia coli:* High-resolution typing method with phylogenetic significance. *Journal of Clinical Microbiology*. **37**: 1274-1279.

Ausubel, F., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. & Struhl K. (1994). *Current Protocols in Molecular Biology*. John Wiley & Sons Inc., New York.

Ball, A. (1999). Annual Review of the Microbiological Quality of Drinking-Water in New Zealand 1998. Ministry of Health, Wellington.

Baudisova, D. (1997). Evaluation of *Escherichia coli* as the main indicator of faecal pollution. *Water Science and Technology*. **35**: 333-36.

Berg, G., Dahling, D.R., Brown, G.A. & Berman D. (1978). Validity of fecal coliforms, total coliforms and fecal streptocci as indicators of viruses in chlorinated primary sewage effluents. *Applied Environmental Microbiology*. **36**: 880-884.

Bernhard, A.E. & Field, K.G. (2000). Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Applied and Environmental Microbiology*. **66**: 1587-1594.

Bettelheim, K.A. (1978). The sources of 'OH' serotypes of *Escherichia coli*. *Journal of Hygiene, Cambridge*. **80**: 83-113.

Bettelheim, K.A., Ismail, N., Shinebaum, R., Shooter, R.A., Moorhouse, E. & Farrell, W. (1976). The distribution of serotypes of *Escherichia coli* in cow-pats and other animal material compared with serotypes of *E. coli* isolated from human sources. *Journal of Hygiene, Cambridge*. **76**: 403-406.

Bielaszewska, M., Schmidt, H., Liesegang, A., Prager, R., Rabsch, W., Tschape, H., Cizek, A., Janda, J., Blahova, K. & Karch, H. (2000). Cattle can be a reservoir of sorbitol-fermenting Shiga toxin-producing *Escherichia coli* O157:H$^-$ strains and a source of human diseases. *Journal of Clinical Microbiology*. **38**: 3470-3473.

Bitton, G. (1994). *Wastewater Microbiology*. Wiley-Liss Inc, New York.

Blears, M.J., de Grandis, S.A., Lee, H. & Trevors, J.T. (1998). Amplified Fragment Length Polymorphism (AFLP): a review of the procedure and its applications. *Journal of Industrial Microbiology and Biotechnology*. **21**: 99-114.

Bonde, G. (1977). Bacterial indication of water pollution. In *Advances in Aquatic Microbiology*. (Droop, M.R. & Jannasch H.W., Eds). pp. 273-362. Academic Press, London.

Brown, C. (1999). Evolution of emerging infections from animals to humans. *The Infectious Disease Review*. **1**: 40-41.

Brown, R.C. & Wade, T.L. (1984). Sedimentary coprostanol and hydrocarbon distribution adjacent to a sewage outfall. *Water Research*. **18**: 621-632.

Cabelli, V.J. (1989). Swimming-associated illness and recreational water quality criteria. *Water Science and Technology*. **21**: 13-21.

Cavé, H., Bingen, E., Elion, J. & Denamur, E. (1994). Differentiation of *Escherichia coli* strains using randomly polymorphic DNA analysis. *Research in Microbiology.* **145**: 141-150.

Chachaty E., Saulnier P., Martin A., Mario N. & Andremont A. (1994). Comparison of ribotyping, pulsed-field gel electrophoresis and Random Amplified Polymorphic DNA for typing *Clostridium difficile* strains. *FEMS Microbiology Letters.* **122**: 61-68.

Clerc, A., Manceau, C. & Nesme, X. (1998). Comparison of Randomly Amplified Polymorphic DNA with Amplified Fragment Length Polymorphism to assess genetic diversity and genetic relatedness within genospecies III of *Pseudomonas syringae*. *Applied Environmental Microbiology.* **64**: 1180-1187.

Close, M.E., Hodgson, L.R. & Tod, G. (1989). Field evaluation of fluorescent whitening agents and sodium tripolyphosphate as indicators of septic tank contamination in domestic wells. *New Zealand Journal of Marine and Freshwater Research.* **23**: 563-568.

Cornax, R., Moriñigo, M., Balebona, C., Castro, D. & Borrego J. (1991). Significance of several bacteriophage groups as indicators of sewage pollution in marine waters. *Water Research.* **25**: 673-678.

Crabill, C., Keim, P. & Southam, G. (1998). Use of AFLP to distinguish human and animal fecal pollution sources in the watershed of Oak Creek, Arizona. In *98th General Meeting of the American Society for Microbiology, Northern Ariz. Univ., Flagstaff, AZ USA, 1998, Abstracts of the General Meeting of the American Society for Microbiology* (American Society for Microbiology), pp. 459. Northern Ariz. Univ., Flagstaff, AZ USA

Crichton, P.B. & Old D.C. (1982). A biotyping scheme for the subspecific discrimination of *Escherchia coli. Journal of Medical Microbiology.* **15**: 233-242.

Davies, C.M., Long, J.A.H., Donald, M. & Ashbolt, N.J. (1995). Survival of fecal microorganisms in marine and freshwater sediments. *Applied and Environmental Microbiology.* **61**: 1888-1896.

Davies-Colley, R.J., Bell, R.G. & Donnison, A.M. (1994). Sunlight inactivation of enterococci and fecal coliforms in sewage effluent diluted in seawater. *Applied and Environmental Microbiology*. **60**: 2049-2058.

Davies-Colley, R.J., Donnison, A.M., Speed, D.J., Ross, C.M. & Nagels, J.W. (1999). Inactivation of faecal indicator microorganisms in waste stabilisation ponds: Interactions of environmental factors with sunlight. *Water Research*. **33**: 1220-1230.

Davis, J.G. (1971). Microbial aspects of pollution: Some general considerations. In *Microbial Aspects of Pollution* (Sykes, G. & Skinner, F.A., Eds), pp. 1-10. Academic Press, London.

de Lamballerie, X., Zandotti, C., Vignoli, C., Bollet, C. & de Mico, P. (1992). A one step microbial DNA extraction method using "Chelex 100" suitable for gene amplification. *Research Microbiology*. **143**: 785-790.

Desai, M., Tanna, A., Wall, R., Efstratiou, A., George, R. & Stanley, J. (1998). Fluorescent Amplified Fragment Length Polymorphism analysis of an outbreak of group a streptococcal invasive disease. *Journal of Clinical Microbiology.* **36**: 3133-3137.

Dijkshoorn, L., Aucken, H., Gernersmidt, P., Janssen, P., Kaufmann, M.E., Garaizar, J., Ursing, J. & Pitt, T.L. (1996). Comparison of outbreak and non-outbreak *Acinetobacter baumannii* strains by genotypic and phenotypic methods. *Journal of Clinical Microbiology.* **34**: 1519-1525.

Dombek, P.E., Johnson, L.K., Zimmerley, S.T. & Sadowsky, M.J. (2000). Use of repetitive DNA sequences and PCR to differentiate *Escherichia coli* isolates from human and animal sources. *Applied and Environmental Microbiology.* **66**: 2572-2577.

Donnison, A.M. & Ross, C.M. (1995). Somatic and F-specific coliphages in New Zealand waste treatment lagoons. *Water Research.* **29**: 1105-1110.

Duim, B., Ang, C.W., van Belkum, A., Rigter, A., van Leeuwen, N.W. J., Endtz, H.P. & Wagenaar, J.A. (2000). Amplified Fragment Length Polymorphism analysis of *Campylobacter jejuni* strains isolated from chickens and from patients with gastroenteritis or Guillain-barre or Miller Fisher syndrome. *Applied and Environmental Microbiology.* **66**: 3917-3923.

Duim, B., Wassenaar, T.M., Rigter, A. & Wagenaar, J. (1999). High-resolution genotyping of *Campylobacter* strains isolated from poultry and humans with Amplified Fragment Length Polymorphism fingerprinting. *Applied Environmental Microbiology.* **65**: 2369-2375.

Duncan, S.H., Botth, I.R., Flint, H.J. & Stewart, C.S. (2000). The potential for the control of *Escherichia coli* O157 in farm animals. *Journal of Applied Microbiology Symposium Supplement.* **88**: 157S-165S.

Edberg, S.C., Le Clerc, H. & Roberston, J. (1997). Natural protection of spring and well drinking waster against surface microbial contamination. II. Indicators and monitoring parameters for parasites. *Critical Reviews in Microbiology.* **23**: 179-206.

Edberg, S.C., Rice, E.W., Karlin, R.J. & Allen, M.J. (2000). *Escherichia coli*: the best biological drinking water indicator for public health protection. *Journal of Applied Microbiology.* **88**: 106S-116S.

Feachem, R. (1975). An improved role for faecal coliform to faecal streptococcal ratios in the differentiation between human and non-human pollution sources. *Water Research.* **9**: 689-690.

Fenlon, D. & Wilson, J. (2000). Growth of *Escherichia coli* O157 in poorly fermented laboratory silage: a possible environmental dimension in the epidemiology of *E. coli* O157. *Letters in Applied Microbiology.* **30**: 118-121.

Fiksdal, L., Maki, J.S., La Croix, S.J. & Staley, J.T. (1985). Survival and detection of *Bacteriodes* spp.: Prospective indicator bacteria. *Applied and Environmental Microbiology.* **49**: 148-150.

Fish, J.T. & Pettibone, G.W. (1995). Influence of freshwater sediment on the survival of *Escherichia coli* and *Salmonella* sp as measured by three methods of enumeration. *Letters in Applied Microbiology.* **20**: 277-281.

Fricker, E.J. & Fricker, C.R. (1994). Application of the polymerase chain reaction to the identification of *Escherichia coli* and coliforms in water. *Letters in Applied Microbiology.* ***19:*** 44-46.

Furuse, K., Ando, A., Osawa, S. & Watanabe, I. (1981). Distribution of ribonucleic acid coliphages in raw sewage from treatment plants in Japan. *Applied and Environmental Microbiology.* **41**: 1139-1143.

Gage, P., Gunther, C.B. & Spaulding, E.H. (1961). Persistence of *Escherichia coli* in the stool samples of infants. *Bacteriological Proceedings.* **61**: 117.

Gibb, A. & Wong, S. (1998). Inhibition of PCR by Agar from bacteriological transport media. *Journal of Clinical Microbiology.* **36**: 275-276.

Gibson, J.R., Slater, E., Xerry, J., Tompkins, D.S. & Owen, R.J. (1998). Use of an Amplified-Fragment Length Polymorphism technique to fingerprint and differentiate isolates of *Helicobacter pylori. Journal of Clinical Microbiology.* **36**: 2580-2585.

Grabow, W.O.K. (1986). Indicator systems for assessment of the virological safety of treated drinking water. *Water Science and Technology.* **18**: 159-165.

Grabow, W.O.K., Coubrough, P., Nupen, E.M. & Bateman, B.W. (1984). Evaluation of coliphages as indicators of the virological quality of sewage-polluted water. *Water South Africa.* **10**: 7-14.

Grabow, W.O.K., Neubrech, T.E., Holtzhausen C.S. & Jofre J. (1995). *Bacteroides fragilis* and *Escherichia coli* bacteriophages - excretion by humans and animals. *Water Science and Technology.* **31**: 223-230.

Grant, M.A., Eklund, C.A. & Shields, S.C. (1995). Monitoring dairy silage for five bacterial groups with potential for human pathogenesis. *Journal of Food Protection.* **58**: 879-883.

Green, D.H., Lewis, G.D., Rodtong, S. & Loutit, M.W. (1991). Detection of faecal pollution in water by an *Escherichia coli uidA* gene probe. *Journal of Microbiological Methods.* **13:** 207-214.

Guinee, P.A.M., Michel, M.F., Agterberg, C.M., Ugueto, N & van Vonno, J. (1972). Resident, transient and uropathogenic *E. coli* strains in the gastrointestinal tract healthy and hospitalised adults. *Antonie van Leenwenhoek,* **38:** 557-566.

Guttman, D.S. (1997). Recombination and clonality in natural populations of *Escherichia coli. Trends in Ecology and Evolution.* **12**: 16-22.

Hagedorn, C., Robinson, S.L., Filtz, J.R., Grubbs, S.M., Angier, T.A. & Reneau, R.B., Jr. (1999). Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. *Applied Environmental Microbiology.* **65**: 5522-5531.

Handzel, T.R., Green, R.M., Sanchez, C., Chung, H. & Sobsey, M.D. (1993). Improved specificity in detecting F-specific coliphages in environmental samples by suppression of somatic phages. *Water Science and Technology.* **27**: 123-131.

Havelaar, A.H., Pot-Hogeboom, W.M., Furuse, K., Pot, R. & Hormann, M.P. (1990). F-specific RNA bacteriophages and sensitive host strains in faeces and wastewater of human and animal origin. *Journal of Applied Bacteriology.* **69**: 30-37.

Hartl, D.L. & Dykhuizen, D.E. (1984). The population genetics of *Escherichia coli*. *Annual Review of Genetics.* **18**: 31-68.

Hartley, C.L., Clements, H.M. & Linton K.B. (1977). *Escherichia coli* in the faecal flora of man. *Journal of Applied Bacteriology*. **43**: 261-269.

Harwood, V.J., Whitlock, J. & Withington, V. (2000). Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: Use in predicting the source of fecal contamination in subtropical waters. *Applied Environmental Microbiology.* **66**: 3698-3704.

Havelaar, A.H., Furuse, K. & Hogeboom, W.M. (1986). Bacteriophage and indicator bacteria in human and animal faeces. *Journal of Applied Bacteriology.* **60**: 255-262.

Havelaar, A.H. & Pot-Hogeboom, W.M. (1988). F-specific RNA bacteriophage as model viruses in water hygiene ecological aspects. *Water, Science and Technology.* **20**: 399-407.

Havelaar, A.H., Pot-Hogeboom, W.M., Furuse, K., Pot, R. & Hormann, M.P. (1990). F-specific RNA bacteriophages and sensitive host strains in faeces and wastewater of human and animal origin. *Journal of Applied Bacteriology.* **69**: 30-37.

Havelaar, A., van Olphen, M. & Drost, Y. (1993). F-specific RNA bacteriophages are adequate model organisms for enteric viruses in fresh water. *Applied Environmental Microbiology.* **59**: 2956-2962.

Herriott, D., Hancock, D., Ebel, E., Carpenter, L., Rice, D. & Besser, T. (1998). Association of herd management factors with colonization of dairy cattle by Shiga toxin-positive *Escherichia coli* O157. *Journal of Food Protection.* **61**: 802-807.

Hinton, M. (1985). The sub-specifc differentiation of *Escherichia coli* with particular reference to ecological studies in young animals including man. *Journal of Hygiene, Cambridge.* **95**: 595-609.

Holdeman, L.V., Good, I.J. & Moore, W.E.C. (1976). Human fecal flora: variation in bacterial composition within individuals and a possible effect of emotional stress. *Applied and Environmental Microbiology.* **31**: 359-375.

Hooper, P. (1999). Viruses from fruit bats to horses, humans and pigs. *The Infectious Diseases Review.* **1**: 214.

Horizons Manawatu-Wanganui. (1999). *Measures of a Changing Landscape; State of the Environment Report Manawatu-Wanganui Region 1999.* D. Young & J McNeill (Eds). Horizons Manawatu-Wanganui, Palmerston North.

Howell, J.M., Coyne, M.S. & Cornelius, P. (1995). Fecal bacteria in agricultural waters of the bluegrass region of Kentucky. *Journal of Environmental Quality.* **24**: 411-419.

Hsu, F.C., Shieh, Y.S.C., Vanduin, J., Beekwilder, M.J & Sobsey, M.D. (1995). Genotyping male-specific RNA coliphages by hybridization with oligonucleotide probes. *Applied and Environmental Microbiology.* **61**: 3960-3966.

IAWPRC. (1991). Review paper: Bacteriophage as model viruses in water quality control. *Water Research.* **25**: 529-545.

Ionas, G., Farrant, K.J., McLenachan, P.A., Clarke, J.K. & Brown, T.J. (1997). Molecular (PCR) differentiation of *Giardia muris* and *Giardia intestinalis. International Journal of Environmental Health Research.* **7**: 63-69.

Jagals, P. & Grabow, W.O.K. (1996). An evaluation of sorbitol-fermenting *Bifidobacteria* as specific indicators of human faecal pollution of environmental water. *Water South Africa.* **22**: 235-238.

Jagals, P., Grabow, W.O.K. & de Villiers J.C. (1995). Evaluation of indicators for assessment of human and animal faecal pollution of surface run-off. *Water Science and Technology.* **31**: 235-241.

Janssen, P., Coopman, R., Huys, G., Swings, J., Bleeker, M., Vos, P., Zabeau, M. & Kersters, K. (1996). Evaluation of the DNA fingerprinting method; AFLP as a new tool in bacterial taxonomy. *Microbiology.* **142**: 1881-1893.

Janssen, P. & Dijkshoorn, L. (1996). High rsolution DNA fingerprinting of *Acinetobacter* outbreak strains. *FEMS Microbiology Letters.* **142**: 191-194.

Jofre, J., Blasi, M., Boasch, A. & Lucena, F. (1989). Occurrence of bacteriophages infecting *Bacteroides fragilis* and other viruses in polluted marine sediments. *Water Science and Technology.* **21**: 15-19.

Jofre, J., Bosch, A., Lucena, F., Gironés, R. & Tartera, C. (1986). Evaluation of *Bacteriodes fragilis* bacteriophages as indicators of virological quality of water. *Water, Science and Technology.* **18**: 167-173.

Johnson, J.R., O'Bryan, T.T., Low, D.A., Ling, G., Delavari, P., Fasching, C., Russo, T.A., Carlino, U. & Stell, A.L. (2000). Evidence of commonality between canine and human extraintestinal pathogenic *Escherichia coli* strains that express papG allele III. *Infection and Immunity.* **68**: 3327-3336.

Joshi, A.K., Baichwal, V. & Ames, G.F. (1991). Rapid Polymerase Chain Reaction amplification using intact bacterial cells. *Biotechniques.* **10**: 44-45.

Kai, M., Watanabe, S., Furuse, K. & Ozawa (1985). *Bacteroides* bacteriophages isolated from human feces. *Microbiology and Immunology.* **29**: 895-899.

Kameyama, L., Fernandez, L., Calderon, J., Ortiz-Rojas, A. & Patterson. T.A. (1999). Characterisation of wild lambdoid bacteriophages: detection of a wide distribution of phage immunity groups and identification of a nus-dependent, non-lambdoid phage group. *Virology*. **263**: 100-111.

Koeleman, J.G.M., Stoof, J., Biesmans, D.J., Savelkoul, P.H.M. & Vandenbroucke-Grauls, C.M.J.E. (1998). Comparison of Amplified Ribosomal DNA Restriction analysis, Random Amplified Polymorphic DNA analysis and Amplified Fragment Length Polymorphism fingerprinting for identification of *Acinetobacter* genomic species and typing of *Acinetobacter baumannii. Journal of Clinical Microbiology.* **36**: 2522-2529.

Kreader, C.A. (1995). Design and evaluation of *Bacteroides* DNA probes for the specific detection of human fecal pollution. *Applied and Environmental Microbiology*. **61**: 1171-1179.

Kreader, C.A. (1996). Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Applied and Environmental Microbiology*. **62**: 1102-1106.

Kreader, C.A. (1998). Persistence of PCR-detectable *Bacteroides distasonis* from human feces in river water. *Applied Environmental Microbiology*. **64**: 4103-4105.

Krumperman, P.H. (1983). Multiple antibiotic resistance indexing of *Escherichia coli* to identify high-risk sources of faecal contamination of foods. *Applied and Environmental Microbiology*. **46**: 165-170.

Kudva, I.T., Hunt, C.W., Williams, C.J., Nance, U.M. & Hovde, C.J. (1997). Evaluation of dietary influences on *Escherichia coli* O157:H7 shedding by sheep. *Applied and Environmental Microbiology*. **63**: 3878-3886.

Kühn, I., Albert, M.J., Ansaruzzaman, M., Bhuiyan, N.A., Alabi, S.A., Islam, M.S., Neogi, P.K.B., Huys, G., Janssen, P., Kersters, K. & Möllby, R. (1997a). Characterization of *Aeromonas* spp. isolated from humans with diarrhea, from healthy controls and from surface water in Bangladesh. *Journal of Clinical Microbiology.* **35**: 369-373.

Kühn, I., Allestam, G., Stenstrom, T.A. & Möllby, R. (1991). Biochemical fingerprinting of water coliform bacteria, a new method for measuring phenotypic diversity and for comparing different bacterial populations. *Applied and Environmental Microbiology.* **57**: 3171-3177.

Kühn, I., Allestam, G., Engdahl, M. & Stenstrom, T.A. (1997b). Biochemical fingerprinting of coliform bacterial populations - comparisons between polluted river water and factory effluents. *Water Science and Technology.* **35**: 343-350.

Kühn, I., Huys, G., Coopman, R., Kersters, K. & Janssen, P. (1997c). A 4-year study of the diversity and persistence of coliforms and *Aeromonas* in the water of a Swedish drinking water well. *Canadian Journal of Microbiology.* **43**: 9-16.

Leeming, R., Ball, A., Ashbolt, N. & Nichols, P. (1996). Using faecal sterols from humans and animals to distinguish faecal pollution in receiving waters. *Water Research.* **30**: 2893-2900.

Leeming, R., Bate, N., Hewlett, R. & Nichols, P.D. (1998). Discriminating faecal pollution: A case study of stormwater entering Port Phillip Bay, Australia. *Water Science and Technology.* **38**: 15-22.

Lees, D.N., Henshilwood, K. & Butcher, S. (1995). Development of a PCR-based method for the detection of enteroviruses and hepatitis A virus in molluscan shellfish and its application to polluted field samples. *Water, Science and Technology.* **31**: 457-464.

Lewis G.D., Molloy S.L., Greening G.E. & Dawson J. (2000). Influence of environmental factors on virus detection by RT-PCR and cell culture. *Journal of Applied Microbiology*. **88**: 633-640.

Linton, A.H. (1977). Animal to man transmission of Enterobacteriaceae. *Royal Society of Health Journal*. **97**: 115-118.

Lipman, L.J., de Nijs, A., Lam, T.J. & Gaastra, W. (1995). Identification of *Escherichia coli* strains from cows with clinical mastitis by serotyping and DNA polymorphism patterns with REP and ERIC primers. *Veterinary Microbiology*. **43**: 13-19.

Lynn, T.V., Hancock, D.D., Besser, T.E., Harrison, J.H., Rice, D.H., Stewart, N.T. & Rowan, L.L. (1998). The occurrence and replication of *Escherichia coli* in cattle feeds. *Journal of Dairy Science*. **81**: 1102-1108.

Madden, R.H., Moran, L. & Scates, P. (1996). Sub-typing of animal and human *Campylobacter* spp. using RAPD. *Letters in Applied Microbiology*. **23**: 167-170.

Mahbubani, M., Bej, A., Perlin, M., Schaefer, III F., Jakubowski, W. & Atlas R. (1991). Detection of *Giardia* cysts by using the Polymerase Chain Reaction and distinguishing live from dead cysts. *Applied and Environmental Microbiology*. **57**: 3456-3461.

Mara, D.D. & Oragui, J.I. (1983). Sorbitol-fermenting *Bifidobacteria. The Journal of Applied Bacteriology*. **55**: 349-358.

Mara, D.D. & Oragui, J.I. (1981). Occurrence of *Rhodococcus coprophilus* and associated actinomycetes in feces, sewage and freshwater. *Applied and Environmental Microbiology*. **42**: 1037-1042.

Mason, T.G. & Richardson, G. (1981). A review *Escherichia coli* and the human gut: Some ecological considerations. *Journal of Applied Bacteriology*. **1**: 1-16.

McFeters, G., Bissonette, G., Jezeski, J., Thomson, C. & Stuart, D. (1974). Comparative survival of indicator bacteria and enteric pathogens in well water. *Applied Microbiology.* **27**: 823-829.

McFeters, G. & Singh, A. (1991). Effects of aquatic environmental stress on enteric bacterial pathogens. *The Journal of Applied Bacteriology Symposium Supplement.* **70**: 115S-120S.

Miao, E.A. & Miller, S.I. (1999). Bacteriophage in the evolution of pathogen-host interactions. *Proceedings of the National Academy of Sciences, USA.* **96**: 9452-9454.

Milkman, R. & McKane, M. (1995). DNA sequence variation and recombination in *E. coli*. In *Population Genetics of Bacteria*. (Baumberg, S., Young, J.P.W., Wellington, E.M.H. & Saunders, J.R., Eds). vol 52, pp127-142. Cambridge University Press, Cambridge.

Milkman, R. & Stoltzfus, A. (1988). Molecular evolution of the *Escherichia coli* chromosome. II Clonal segments. *Genetics.* **120**: 359-366.

Ministry for the Environment. (1997). *Draft-Fresh Water Microbiological Standards and Guideline Programme; Project Plan*. Ministry for the Environment, Wellington.

Ministry for the Environment. (1999a). *Recreational Water Guidelines*. Ministry for the Environment, Wellington.

Ministry for the Environment. (1999b). *Supporting Manual to the 1998 Bacteriological Water Quality Guidelines for Marine and Fresh Water.* Ministry for the Environment, Wellington.

Ministry of Health. (1995). *Guidelines on Drinking-Water Quality Management for New Zealand.* Ministry of Health, Wellington.

Ministry of Health. (2000). *Drinking Water Standards for New Zealand.* Ministry of Health, Wellington.

Moore, W.E.C. & Holdeman, L.V. (1974). Human faecal flora: The normal flora of 20 Japanese-Hawaiians. *Applied Microbiology.* **27**: 961-979.

Moshetti, G., Blaiotta, G., Aponte, M., Catzeddu, P. Villani F., Deiana P. & Coppola, S. (1998). Random Amplified Polymorphic DNA and Amplified Ribosomal DNA Spacer Polymorphism powerful methods to differentiate *Streptococcus thermophilus* strains. *Journal of Applied Microbiology.* **85**: 25-36.

Mueller, U.G. & Wolfenbarger, L.L. (1999). AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution.* **14**: 389-394.

Niemi, M., Sibakov, M. & Niemela, S. (1983). Antibiotic resistance among different species of fecal coliforms isolated from water samples. *Applied and Environmental Microbiology.* **45**: 79-83.

Nix, P.G., Daykin, M.M. & Vilkas, K.L. (1994). Fecal pollution events reconstructed and sources identified using a sediment bag grid. *Water Environment Research.* **66**: 814-818.

Obiri-Danso, K. & Jones, K. (1999). The effect of a new sewage treatment plant on faecal indicator numbers, Campylobacters and bathing water compliance in Morecambe Bay. *Journal of Applied Microbiology.* **86**: 603-614.

Ochman, H., Whittam, T., Caugant, D.A. & Selander, R. (1983). Enzyme polymorphism and genetic population in *Escherichia coli* and *Shigella*. *Journal of General Microbiology.* **129**: 2715-2726.

Olmos, A., Camarena, J.J., Nogueira, J.M., Navarro, J.C., Risen, J. & Sánchez, R. (1998). Application of an optimized and highly discriminatory method based on Arbitrarily Primer PCR for epidemiologic analysis of methicillin-resistant *Staphlococcus aureus* nosocomial infections. *Journal of Clinical Microbiology*. **36**: 1128-1134.

Oragui, J.I. & Mara, D.D. (1983). Investigation of the survival characteristics of *Rhodococcus coprophilus* and certain faecal indicator bacteria. *Applied and Environmental Microbiology*. **46**: 356-360.

Osawa, S., Furuse, K. & Watanabe, I. (1981). Distribution of Ribonucleic Acid coliphage in animals. *Applied and Environmental Microbiology*. **41**: 164-168.

Oshiro, R. & Fujioka, R. (1995). Sand, soil, and pigeon droppings: Sources of indicator bacteria in the waters of Hanauma Bay, Oahu, Hawaii. *Water, Science and Technology*. **31**: 251-254.

Parveen, S., Murphree, R.L., Edmiston, L., Kaspar, C.W., Portier, K.M. & Tamplin M.L. (1997). Association of multiple-antibiotic-resistance profiles with point and nonpoint sources of *Escherichia coli* in Apalachicola Bay. *Applied and Environmental Microbiology*. **63**: 2607-2612.

PE Applied Biosystems. (1996). *AFLP™ Microbial Fingerprinting Protocol.* The Perkin-Elmer Corporation, Foster City.

Isono, K. (1997). A novel quantitative PCR with fluorogenic probe. *Rinsho Byori*. **43**: 218-223.

Petrocheilou, V., Richmond, M.H. & Bennett, P.M. (1979). Persistence of plasmid-carrying tetracycline-resistant *Escherichia coli* in a married couple, one of whom was receiving antibiotics. *Antimicrobial Agents and Chemotherapy*. **16**: 225-230.

Pettibone, G.W., Sullivan, S.A.M. & Shiaris, M.P. (1987). Comparative survival of antibiotic-resistant and -sensitive fecal indicator bacteria in estuarine water. *Applied and Environmental Microbiology.* **53**: 1241-1245.

Power, E.G.M. (1996). RAPD typing in microbiology - a technical review. *Journal of Hospital Infection.* **34**: 247-265.

Puig, A., Jofre, J. & Araujo, R. (1997). Bacteriophages infecting various *Bacteroides fragilis* strains differ in their capacity to distinguish human from animal faecal pollution. *Water Science and Technology.* **35**: 359-362.

Puig, A., Queralt, N., Jofre, J. & Araujo, R. (1999). Diversity of *Bacteroides fragilis* strains in their capacity to recover phages from human and animal wastes and from fecally polluted wastewater. *Applied and Environmental Microbiology.* **65**: 1772-1776.

Pupo, G.M. & Richardson, B.J. (1995). Biochemical genetics of a natural population of *Escherichia coli* - seasonal changes in alleles and haplotypes. *Microbiology.* **141**: 1037-1044.

Rabouam, C., Comes, A.M., Bretagnolle, V., Humbert, J.F., Periquet, G. & Bigot, Y. (1999). Features of DNA fragments obtained by Random Amplified Polymorphic DNA (RAPD) assays. *Molecular Ecology.* **8**: 493-503.

Reeves, P.R. (1992). Variation in O-antigens, niche-specific selection and bacterial populations. *FEMS Microbiology Letters.* **100**: 509-516.

Resnick, I.G. & Levin, M.A. (1981). Assessment of *Bifidobacteria* as indicators of human faecal pollution. *Applied and Environmental Microbiology.* **42**: 433-438.

Roll, B.M. & Fujioka, R.S. (1997). Sources of faecal indicator bacteria in a brackish, tropical stream and their impact on recreational water quality. *Water Science and Technology.* **35**: 179- 186.

Rothmaier, R., Weidenmann, A. & Botzenhart, K. (1997). Transport of *Escherichia coli* through soil to groundwater traced by Randomly Amplified Polymorphic DNA (RAPD). *Water Science and Technology.* **35**: 351-357.

Rowbotham, T.J. & Cross, T. (1977). Ecology of *Rhodococcus coprophilus* and associated actinomycetes in freshwater an agricultural habitats. *Journal of General Microbiology.* **100**: 231-240.

Sabat, G., Rose, P., Hickey, W.J. & Harkin, J.M. (2000). Selective and sensitive method for PCR amplification of *Escherichia coli* 16S rRNA genes in soil. *Applied and Environmental Microbiology.* **66**: 844-849.

Sambrook, J. & Russell, D.W. (2001). *Molecular Cloning: A Laboratory Manual.* Volume 1. 3rd Edition. Cold Springer Harvour Laboratory Press, New York. pp 1.44-1.46.

Sears, H.J., Brownlee, I. & Uchiyama, J.H. (1950). Persistence of individual strains of *Escherichia coli* in the intestinal tract of man. *Journal of Bacteriology.* **59**: 293-301.

Sears, H.J., Janes, H., Saloum, R., Brownlee, I. & Lamoreaux, L.F. (1956). Persistence of individual strains of *Escherichia coli* in man and dog under varying conditions. *Journal of Bacteriology.* **71**: 370-372.

Serrano, E., Moreno, B., Solaun, M., Aurrekoetxea, J.J. & Ibarluzea, J. (1998). The influence of environmental factors on microbiological indicators of coastal water pollution. *Water Science and Technology.* **38**: 195-199.

Shadix, L.C. & Rice, E.W. (1991). Evaluation of β-glucuronidase assay for the detection of *Escherichia coli* from environmental waters. *Canadian Journal of Microbiology.* **37**: 908-911.

Sinton, L.W. & Donnison, A.M. (1994). Characterisation of faecal streptococci from some New Zealand effluents and receiving waters. *New Zealand Journal of Marine and Freshwater Research.* **28**: 145-158.

Sinton, L.W., Donnison A.M. & Hastie C.M. (1993a). Faecal streptococci as faecal pollution indicators - A review. Part I: Taxonomy and enumeration. *New Zealand Journal of Marine and Freshwater Research.* **27**: 101-115.

Sinton, L.W., Donnison, A.M. & Hastie, C.M. (1993b). Faecal streptococci as faecal pollution indicators - A review. Part II: Sanitary significance, survival and use. *New Zealand Journal of Marine and Freshwater Research.* **27**: 117-137.

Sinton, L.W. & Finlay, R.K. (1996). Bacteriophages as microbiological water quality indicators. *Water and Wastes in New Zealand.* **March**: 52-55.

Sinton, L.W., Finlay, R.K. & Hannah, D.J. (1998). Distinguishing human from animal faecal contamination in water - A review. *New Zealand Journal of Marine and Freshwater Research.* **32**: 323-348.

Sinton. L.W., Finlay, R.K. & Lynch, P.A. (1999). Sunlight inactivation of fecal bacteriophages and bacteria in sewage-polluted seawater. *Applied and Environmental Microbiology.* **65**: 3605-3613.

Soothill, J. & Mountford J.C. (1999). The prevelance of antibiotic-resistant strains of *Escherichia coli* in the faeces of meat-eaters and vegetarians. *The Infectious Disease Review.* **1**: 32.

Souza, V., Rocha, M., Valera, A. & Eguiarte, L.E. (1999). Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. *Applied Environmental Microbiology.* **65**: 3373-3385.

Speijer, H., Savelkoul, P.H.M., Bonten, M.J., Stobberingh, E.E. & Tjhie, J.H.T. (1999). Application of different genotyping methods for *Pseudomonas aeruginosa* in a setting of endemicity in an intensive care unit. *Journal of Clinical Microbiology.* **37**: 3654-3661.

Swayne, D.E. (1999). Infection of mammals with avian influenze virus. *The Infectious Disease Review.* **1**: 214.

Tartera, C. & Jofre, J. (1987). Bacteriophages active against *Bacteroides fragilis* in sewage-polluted water. *Applied Environmental Microbiology.* **53**: 1632-1637.

Tartera, C., Lucena, F. & Jofre, J. (1987). Human origin of *Bacteriodes fragilis* bacteriophage present in the environment. *Applied and Environmental Microbiology.* **55**: 2696-2701.

Taylor, J. (1961). Host specificity and enteropathogenicity of *Escherichia coli. Journal of Applied Bacteriology.* **24**: 316-325.

Tilsala-Timisjarvi, A. & Alatossava, T. (1998). Strain-specific identification of probiotic *Lactobacillus rhamnosus* with Randomly Amplified Polymorphic DNA-derived PCR primers. *Applied and Environmental Microbiology.* **64**: 4816-4819.

Tsai, Y.L., Palmer, C.J. & Sangermano, L.R. (1993). Detection of *Escherichia coli* in sewage and sludge by Polymerase Chain Reaction. *Applied and Environmental Microbiology.* **59**: 353-357.

Turner, S.J. (1996). *Microbiological Indicators of Faecal Contamination in Aquatic Environments.* PhD Thesis. School of Biological Sciences, University of Auckland, Auckland, New Zealand.

Turner, S.J., Lewis, G.D. & Bellamy, A.R. (1997a). A genomic polymorphism located downstream of the *gcv*P gene of *Escherichia coli* that correlates with ecological niche. *Molecular Ecology.* **6**: 1019-1032.

Turner, S.J, Lewis, G.D. & Bellamy, A.R. (1997b). Detection of sewage-derived *Escherichia coli* in a rural stream using multiplex PCR and automated DNA detection. *Water Science and Technology.* **35**: 337-342.

Valsangiacomo, C., Baggi, F., Gaia, V., Balmelli, T., Peduzzi, R. & Piffaretti, J.C. (1995). Use of Amplified Fragment Length Polymorphism in molecular typing of *Legionella pneumophila* and application to epidemiological studies. *Journal of Clinical Microbiology.* **33**: 1716-1719.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Horners, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. & Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucelic Acids Research.* **23**: 4407-4414.

Wang, G. Whittam T.S. Berg, C.M. & Berg, D.E. (1993). RAPD (arbitrary primer) PCR is more sensitive than multilocus enzyme electrophoresis for distinguishing related bacterial strains. *Nucleic Acids Research.* **21**: 5930-5933.

Warner, J.M & Oliver, J.D. (1999). Randomly Amplified Polymorphic DNA analysis of clinical and environmental isolates of *Vibrio vulnificus* and other *Vibrio* species. *Applied and Environmental Microbiology.* **65**: 1141-1144.

Welsh, J. & McClelland. M. (1991). Genomic fingerprinting using arbitrarily primed PCR and a matrix of pairwise combinations of primers. *Nucleic Acids Research.* **19**: 5275-5279.

Welsh, J. & McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research.* **18**: 7213-7218.

West, P.A. (1991). Human pathogenic viruses and parasites: emerging pathogens in the water cycle. *Journal of Applied Bacteriology Symposium Supplement.* **70**: 107S-114S.

Wiggins, B.A. (1996). Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters. *Applied and Environmental Microbiology.* **62**: 3997-4002.

Wiggins, B.A., Andrews, R.W., Conway, R.A., Corr, C.L., Dobratz, E.J., Dougherty, D.P., Eppard, J.R., Knupp, S.R., Limjoco, M.C., Mettenburg, J.M., Rinehardt, J.M., Sonsino, J., Torrijos, R.L. & Zimmerman, M.E. (1999). Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. *Applied and Environmental Microbiology.* **65**: 3483-3486.

Wilcock, R.J. (1986). Agricultural run-off: A source of water pollution in New Zealand? *New Zealand Journal of Agricultural Science.* **20**: 98-103.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A.& Tingey, S.V. (1990). DNA polymorphisms amplified by arbitary primers are useful as genetic markers. *Nucleic Acids Research.* **18**: 6531-6535.

Zhao, S., Mitchell, S.E., Meng, J., Kresovich, S., Doyle, M.P., Dean, R.E.C.A.M. & Weller, J.W. (2000). Genomic typing of *Escherichia coli* O157:H7 by semi-automated fluorescent AFLP analysis. *Microbes and Infection.* **2**: 107-113.

APPENDIX A - REAGENTS

A.1 1x E buffer

62.5mM Tris, 1mM EDTA-disodium salt, 0.5mM sodium acetate, pH 7.8.

A.2 1 x TE buffer

TE buffer comprised 10 mM Tris-HCl, pH 8.0, containing 1 mM ETDA.

A.3 Loading Dye for Agarose Gels

0.2% w/v bromophenol blue, 0.25% w/v xylene cyanol, 15% w/v ficoll, 1 x E-buffer.

A.4 Formamide Loading Dye for Polyacrylamide Gels

98% v/v formamide, 10 mM 0.5M EDTA, pH 8.0, 0.05% w/v bromophenol blue, 0.05% w/v xylene cyanol.

A.5 AFLP Adapter Preparation

Eco RI Adapter

The *Eco* RI Adapter was prepared as follows. The oligonucleotides, *Eco* RI-1 and *Eco* RI-2 were prepared (see below) with 1 x TE buffer pH 8.3 to a final concentration of 5 pmole/μL. The samples were heated to 95°C for 5 min, and then slowly returned to room temperature.

Mse I Adapter

The Mse I adapter was prepared in the same manner, but to a final concentration of 50 pmol/μL.

When combined, the oligonucleotides anneal as follows to form the adpater.

Eco RI-1	5'	CTC GTA GAC TGC GTA CC	3'
Eco RI-2	3'	CAT CTG ACG CAT GGTTAA	5'

Mse I-1	5'	GAC GAT GAG TCC TGA G	3'
Mse I-2	3'	TA CTC AGG ACT CAT	5'

APPENDIX B – PCR OPTIMISATION OF THE DAIRY FAECAL MARKER

The affect of various parameters on the efficiency of target-specific PCR (TS-PCR) amplification of the dairy faecal marker was investigated, including primer annealing temperature, $MgCl_2$ concentration and number of PCR amplification cycles.

B.1 $MgCl_2$ Optimisation

B.1.1 Method

The effect of $MgCl_2$ concentration on amplification efficiency was investigated as follows. Reactions contained 2 μL of 10 x *Taq* polymerase buffer (minus $MgCl_2$) (Gibco, BRL), 10 pmole each primer (Dairy 1B and Dairy 2B; Table 4.1), 200 μM each dNTP (dATP, dCTP, dGTP, dTTP) (Roche Diagnostics), 1.0 U *Taq* polymerase and 1.0 μL DNA lysate (Section 4.3.2). $MgCl_2$ was added at the following concentrations — 0.75, 1.0, 1.25, 1.5, 1.75, 2.0 and 2.5 mM. Sterile Milli-Q water was added to a final volume of 20 μL. Negative controls were included in all experiments. Two isolates from dairy oxidation ponds, two human isolates and one isolate from a sewage pond were analysed.

The amplification program comprised of 1 cycle of 94°C for 30 sec; 25 cycles of 94°C for 30 sec, 59°C for 30 sec, 72°C and 30 sec; 1 cycle of 94°C for 30 sec, 59°C for 30 sec, and 72°C for 5 min and held at 4°C until required. Target-specific PCR products were separated by gel electrophoresis on 1.6% agarose, and were visualised by UV illumination as described in Section (Section 4.2.2.4). 1 Kb Plus DNA Ladder™ was electrophoresed in adjacent lanes to determine product size.

B.1.2 Results & Conclusions

Amplification of the dairy faecal indicator (462 bp) was obtained at all concentrations of $MgCl_2$. There appeared little difference in level of amplification of

the dairy marker when the $MgCl_2$ concentration was between 0.75 mM and 1.5 mM (Figure B.1.A, Lanes 1-18). At concentrations greater than 1.75 mM, the level of amplified target product increased (Figure B.1, Gel A, Lanes 19-20 versus Gel B Lanes 1-2), however this may have largely been a difference in gel staining. Unfortunately, the occurrence of several non-specific PCR products (400-850 bp) in the human isolates also increased (Gel A, Lane 21 versus Gel B, Lane 3), which although barely visible in the reproductions shown here, were more clearly defined in the original image.

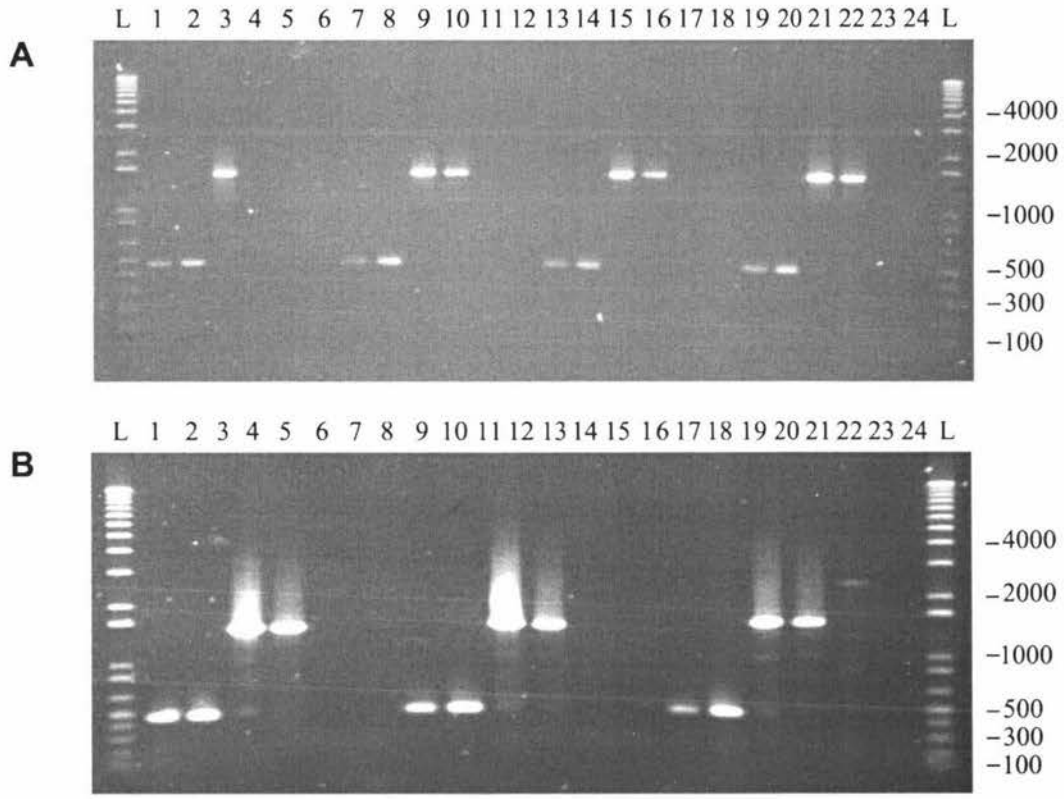


Figure B.1. Optimisation of $MgCl_2$ concentration for TS-PCR amplification of the dairy marker. Lane L, 1 Kb Plus DNA Ladder™; **(A)** Lanes 1 - 6, 0.75 mM $MgCl_2$; Lanes 7 – 12, 1.0 mM $MgCl_2$; Lanes 13 – 18, 1.25 mM $MgCl_2$; Lanes 19 – 24, 1.5 mM $MgCl_2$; **(B)** Lanes 1 - 6, 1.75 mM $MgCl_2$; Lanes 7 – 12, 2.0 mM $MgCl_2$; Lanes 13 – 18, 2.5 mM $MgCl_2$; **(A & B)** Lanes 1, 7, 13 & 19, dairy pond isolate D1/1; Lanes 2, 8, 14 & 20, dairy pond isolate D2/2; Lanes 3, 9, 15 & 21, sewage isolate HS1/1; Lanes 4, 10, 16 & 22, human isolate 71/1; Lanes 5, 11, 17 & 23, human isolate 43/3; Lanes 6, 12, 18 & 24, negative control.

An additional non-target fragment at 1470 bp (A, Lane 21) was consistently observed from a limited number of human isolates, and could not be eliminated by optimising $MgCl_2$ concentration. This fragment appeared related to the target 462 bp product, as shown by restriction enzyme digestion (Section 4.4.3). Human isolates that possessed the non-target fragment, were specifically included in the optimisation experiments to minimise its amplification. The absence of the 1470 bp product at 0.7 mM $MgCl_2$ in the human isolate (A, Lane 4) may have been due to PCR failure, resulting from pipetting error.

The optimum $MgCl_2$ concentration was approximately 1.5 mM. 1.75 mM may also have been suitable, however the emergence of minority non-specific products and the availibility of pre-mixed commercial preparations of *Taq* polymerase buffer containing 1.5 mM $MgCl_2$, the latter concentration was used for subsequent experiments.

B.2 Primer Annealing Temperature

B.2.1 Method

Optimisation of primer annealing temperature was performed as described in Section B.1.1, using 1.5mM $MgCl_2$ and an annealing temperature gradient. The amplification program comprised of 1 cycle of 94°C for 30 sec; 25 cycles of 94°C for 30 sec, X°C[1] for 30 sec, 72°C for 30 sec; 1 cycle of 94°C for 30 sec, X°C[1] for 30 sec, 72°C for 5 min was and held at 4°C until required. The temperature gradient function of the Eppendorf Mastercycler® was utilised during primer annealing, so that the annealing temperature increased sequentially (59°C ± 5°C) across the block from 54°C to 64°C, enabling simultaneous analysis of several annealing temperatures. Genotypic fingerprints were compared at each temperature for two dairy pond isolates, two human isolates and one isolate from sewage.

[1] Annealing temperature was set at 59°C with a gradient of ±5°C.

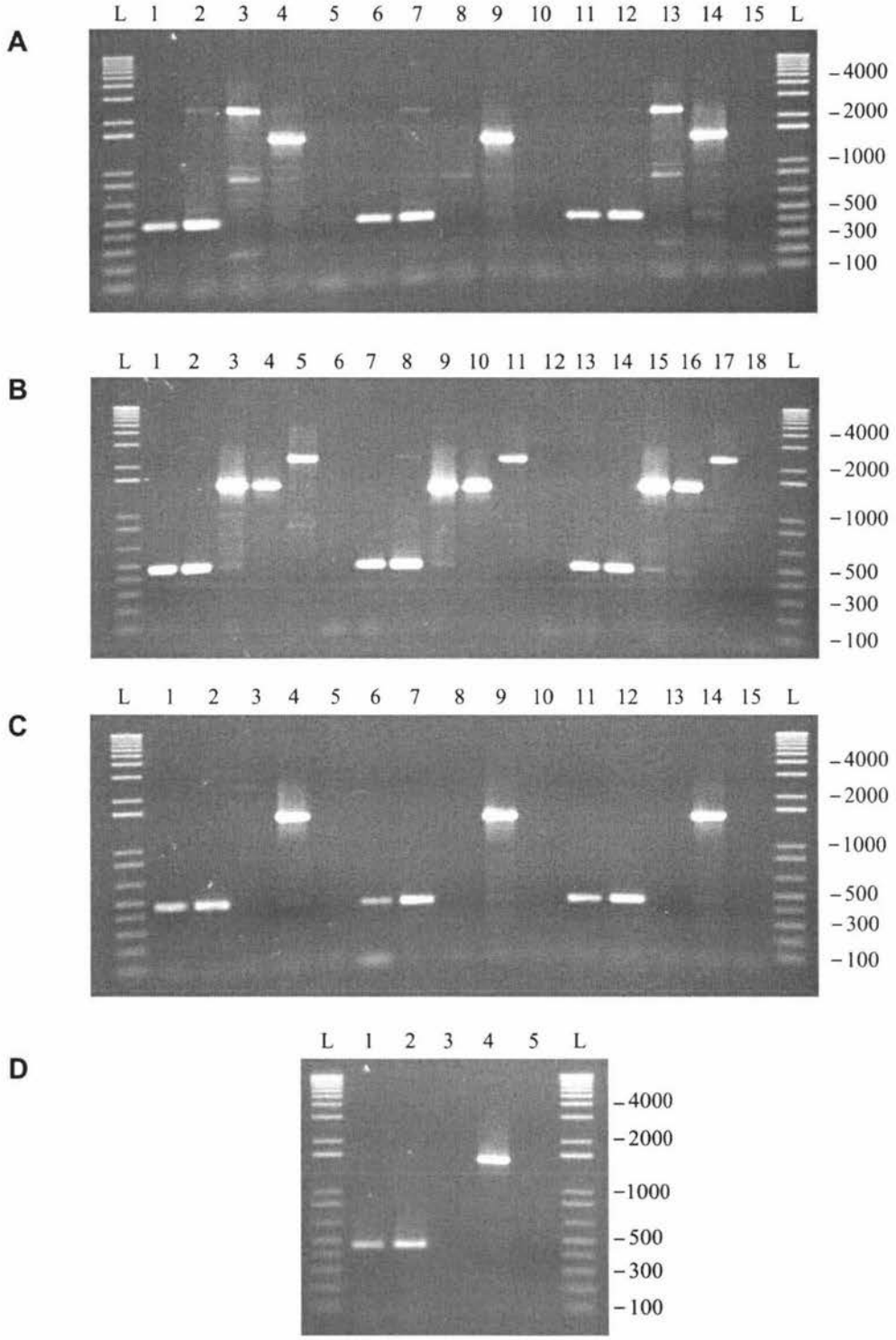


Figure B.2. Optimisation of primer annealing temperature for TS-PCR amplification of the dairy marker. Lane L, 1 Kb Plus DNA Ladder™; **(A)** Lanes 1 - 5, 54.1°C; Lanes 6 – 10, 54.5°C; Lanes 11 – 15, 55.4°C; **(B)** Lanes 1 - 5, 56.6°C; Lanes 6 – 10, 57.9°C; Lanes 11 – 15, 59.4°C; **(C)** Lanes 1 - 5, 60.8°C; Lanes 6 – 10, 62.0°C; Lanes 11 – 15, 63.0°C; **(D)** Lanes 1 - 5, 63.7°C; **(A, B, C & D)** Lanes 1, 6 & 11, dairy pond isolate D1/1; Lanes 2, 7 & 12, dairy pond isolate D2/2; Lanes 3, 8 & 13, human isolate 43/3; Lanes 4, 9 & 14, human isolate 71/1; Lanes 5, 10 & 15, negative control.

B.2.2 Results & Conclusions

Figure B.2 shows that non-target products (*e. g.* 1470 and 2360 bp) were consistently amplified from human isolates at low annealing temperatures (A, Lanes 3 & 4). Several non-target products were also amplified from dairy isolates at these temperatures (A, Lane 2). Although amplification of the fragment at 2360 bp was eliminated at higher annealing temperatures (C, Lane 5, 10, 15,), the band at 1470 bp was not reduced. Amplification efficiency of the dairy marker declined slightly as annealing temperature increased (A, Lane 1 versus C, Lane 6), however the fragment was still clearly defined. An annealing temperature of 62°C, minimised amplification of non-target products but enabled suitable amplification of the target marker. Consequently, this temperature was utilised for further analyses.

B.3 Number of PCR Cycles

B.3.1 Method

Optimisation of the number of PCR cycles was performed as described in Section B.1.1, using 1.5mM $MgCl_2$, 250μM each dNTP (dATP, dCTP, dGTP, dTTP) and an annealing temperature of 62°C. PCR profiles from four strains (two dairy and two human) were compared after various cycles of amplification were completed (10, 15, 20 and 25).

B.3.2 Results & Conclusions

Twenty-five PCR amplification cycles provided optimum amplification of the diagnostic dairy fragment (Figure B.3; Lanes 16- 17). Although thirty or thirty-five cycles could have been investigated as well, the diagnostic marker was definitively detected with the minimum analysis time.

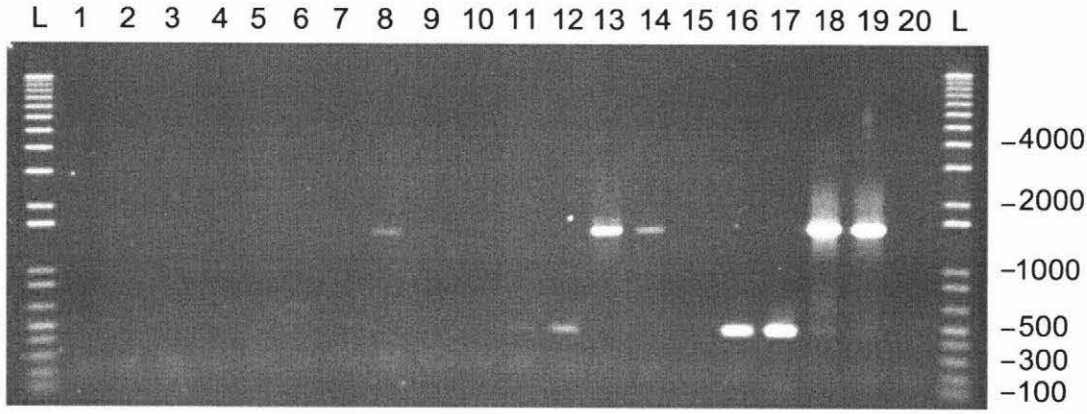


Figure B.3. Optimisation of the number of amplification cycles for TS-PCR amplification of the dairy marker. Lane L, 1 Kb Plus DNA Ladder™; Lanes 1 - 5, 10 cycles; Lanes 6 – 10, 15 cycles; Lanes 11 – 15, 20 cycles; Lanes 16 – 20, 25 cycles; Lanes 1,6, 11 & 16, dairy pond isolate D1/1; Lanes 2, 7, 12 & 17, dairy pond isolate D2/2; Lanes 3, 8, 13 & 18, sewage isolate HS1/1; Lanes 4, 9, 14 & 19, human isolate 71/1; Lanes 5, 10, 15 & 20, negative control.

B.4 Taq Polymerase Brand

B.4.1 Method

Several brands of *Taq* polymerase were compared to verify that amplification was not affected by different brands of polymerase. The comparison was undertaken as described in Section B.1.1, using 1.5 mM $MgCl_2$, 250 μM of each dNTP, an annealing temperature of 62°C and twenty-five PCR amplification cycles. The following *Taq* polymerases were analysed – *Taq* polymerase (Roche Diagnostics), MasterTaq (Eppendorf) and Platinum *Taq* (Gibco BRL). The effect on amplification efficiency of each of the enzymes was compared using six dairy pond isolates and one isolate from sewage.

B.5.2 Results & Conclusions

Figure B.4 shows that there was little variation in amplification efficiency, which could be attributed to different brands of *Taq* polymerase. Consequently, all three brands were suitable for use, and *Taq* polymerase from Roche Diagnostics was used in subsequent experiments.

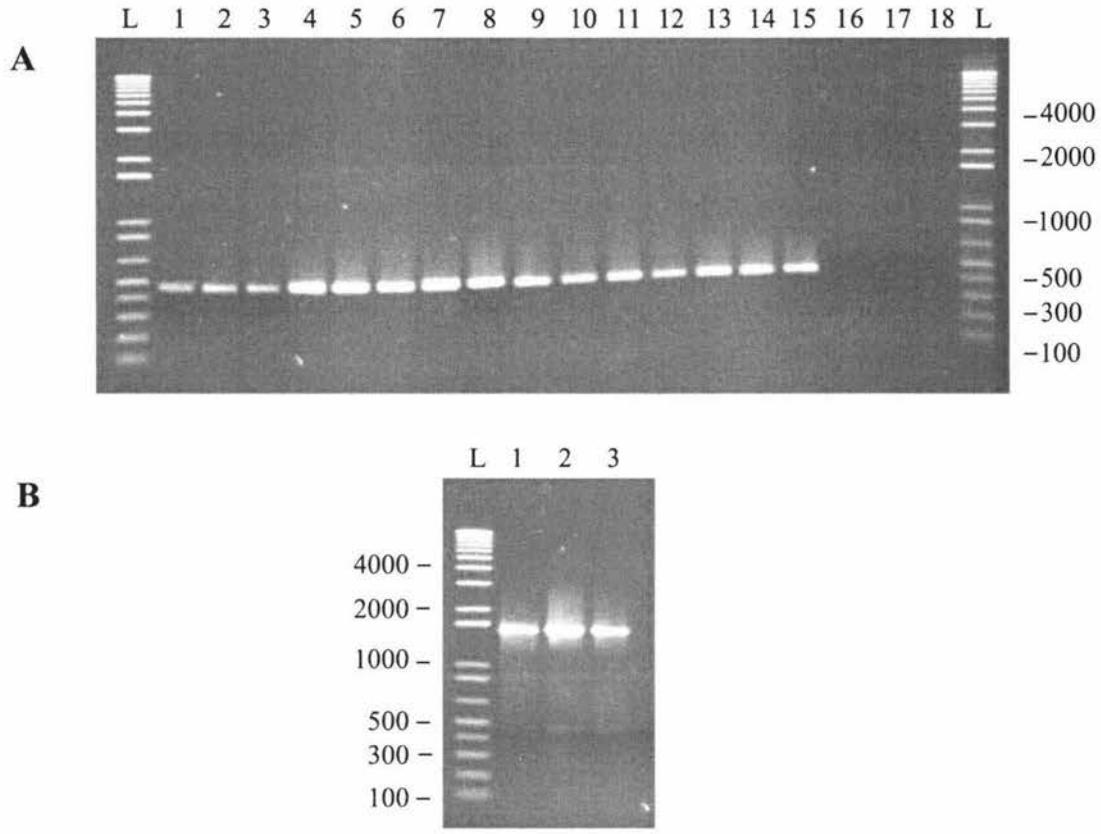


Figure B.4. Comparison of *Taq* polymerase brand for TS-PCR amplification of the dairy marker. Lane L, 1 Kb Plus DNA Ladder™; **(A & B)** Lanes 1, 4, 7, 10, 13 & 16, *Taq* polymerase; Lanes 2, 5, 8, 11, 14 & 17, MasterTaq; Lanes 3, 6, 9, 12, 15 & 18 Platinum *Taq*; **(A)** Lanes 1 - 3, dairy pond isolate D1/2; Lanes 4-6, dairy pond isolate D2/3; Lanes 7 – 9, dairy pond isolate D2/4; Lanes 10-12, dairy pond isolate D3/3; Lanes 13 – 15, dairy pond isolate D3/10; Lanes 16 – 18 dairy pond isolate D3/11; **(B)** Lanes 1 - 3, HS1/1; Lanes 4-6, human isolates 71/1; Lanes 7 – 9, human isolate 43/3.

B.5 Summary - Optimised TS-PCR Conditions for Dairy Faecal Marker

The optimised conditions for the putative dairy faecal marker were as follows. Reactions contained 2 µL 10 x *Taq* polymerase buffer, 10 pmole each primer (Dairy 1B and Dairy 2B, Table 5.1), 250 µM each dNTP (dATP, dCTP, dGTP, dTTP) (Roche Diagnostics), 1.0 U *Taq* polymerase and 1.0 µL DNA lysate (Section 4.3.2). Sterile Milli-Q water was added to a final volume of 20 µL. Positive and negative controls were included in all experiments. The amplification program comprised of 1 cycle of 94°C for 30 sec; 25 cycles of 94°C for 30 sec, 62°C for 30 sec, 72°C for 30 sec; 1 cycle of 94°C for 30 sec, 62°C for 30 sec, 72°C for 5 min and held at 4°C until required. PCR amplification was performed using an Eppendorf Mastercycler®.

APPENDIX C

C.1 pGEM®-T EASY VECTOR SEQUENCE

pUC/M13 Forward Primer

2951 CCAGGGTTTT CCCAGTCACG ACGTTGTAAA ACGACGGCCA GTGAATTGTA

ATACGACTCA CTATA GGGCGAATTG GGCCCGACGT CGCATGCTCC GGCCGCC

Eco RI Eco RI

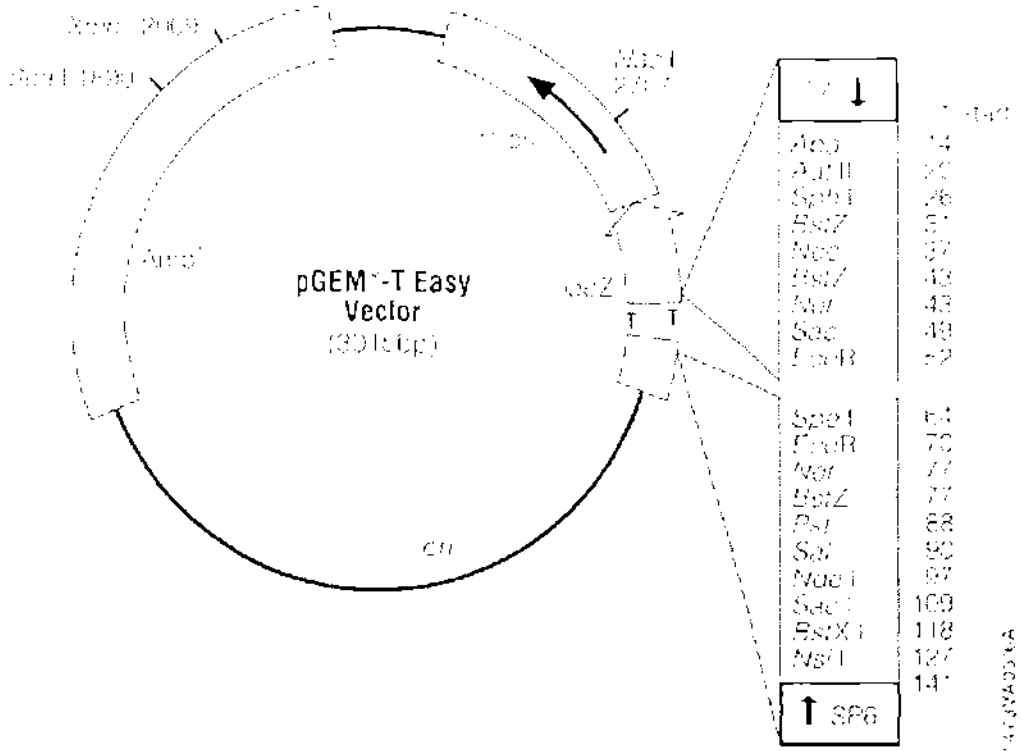
AT GGCGGCCGCG GGAATTCGAT (Insertion) ATCACTAGTG AATTCGCGGC CGC

CTGCAGG TCGACCATAT GGGAGAGCTC CCAACGCGTT GGATGCATAG CTTGA

GTATT CTATAGTGTC ACCTAAATAG CTTGGCGTAA TCATGGTCAT AGCTGTTT

pUC/M13 Reverse Primer

CC TGTGTGAAAT 2486



Taken from Promega pGEM® -T Easy Technical Bulletin

C.2 ALIGNMENT OF DAIRY POLYMORPHISM WITH *E. COLI* K12

Polymorphic fragment from dairy isolates (includes the 3 selective base overhang from the AFLP primers) aligned with Section 142 of 400 of the complete genome of *E. coli* K12.

Score = 845 bits (419), Expect = 0.0; Identities = 544/590 (92%), Gaps = 6/590 (1%) Strand = Plus / Minus

```
Dairy Marker:
      1 ctatattgtgaggsytngcataatggcattcagaatgagtgaacaaccmcggrnccataa 60
        |||||||||||||  | |||||||||||||||||||||||||||||||| |||  ||||||
  2521 ctatattgtgaggctt-gcataatggcattcagaatgagtgaacaaccacgga-ccataa 2464
E. coli K12

Dairy Marker:
     61 aaatttataatynntgctggccggranctaatgaatttattggtgaaggtgacgcatata 120
        ||||||||||||   |||||||||| | |||||||||||||||||||||||| |||||||
  2463 aaatttataatc--tgctggccggaa-ctaatgaatttattggtgaaggtgatgcatata 2407
E. coli K12:

Dairy Marker: 121
    121 ttccgcctcataccggtctgcctgcaaacagtaccgatattgcaccgccagatattccgg 180
        |||||||||||| |||||||| ||||||||||||||||||||||||||||||||||||
  2406 ttccgcctcatacaggtctgccagcaaacagtaccgatattgcaccgccagatattccgg 2347
E. coli K12: 2406

Dairy Marker:
    181 ctggctttgtggctgttttcaacagtgatgaggcatcgtggcatctcgttgaagaccatc 240
        ||||||| |||||||||||||||||||| |||||||||||||||||||| ||||
  2346 ctggcttcgtggctgttttcaacagtgatgagtcatcgtggcatctcgttgaagatcatc 2287
E. coli K12:

Dairy Marker:
    241 ggggtaaaaccgtctatgacgtggcttccggcgacgcgttatttatttctgaactcggtc 300
        |||||||||| || ||||||||||||||||||||||||||||||||||||||||||||||
  2286 ggggtaaaacggtttatgacgtggcttccggcgacgcgttatttatttctgaactcggtc 2227
E. coli K12:

Dairy Marker:
    301 cgttaccggaaaattttacctggttatcgccgggaggggaatatcagaagtggaacggca 360
        |||||||||||||| |||||||||||||||||| |||||| | ||||||||||||||||
  2226 cgttaccggaaaatgttacctggttatcgccggaaggggagtttcagaagtggaacggca 2167
E. coli K12:

Dairy Marker:
    361 cagcctgggtgaaggatacggaagcagaaarrctgttccggatccgggaggcggaagaaa 420
        |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
  2166 cagcctgggtgaaggatacggaagcagaaaaactgttccggatccgggaggcggaagaaa 2107
E. coli K12:

Dairy Marker:
    421 cnnnnnnnngcctgatgcaggtagccagtgagcatattgcgnccgcttcaggatgctgca 480
        |          |||||||||||||||||||||||||||||| ||||||||||||||||||
  2106 caaaaaacaacctgatgcaggtagccagtgagcatattgcg-ccgcttcaggatgctgca 2048
E. coli K12:

Dairy Marker:
    481 gatctggaaattgcaacggaggaagaaacctcgttgctggaarcctggaagaagtatcgg 540
        |||||||||||||||||||||||||||||| |||||||||||| | ||||| |||||||||
  2047 gatctggaaattgcaacggaggaagaaatctcgttgctggaagcatggaaaaagtatcgg 1988
E. coli K12:

Dairy Marker:
    541 gtattgctgaaccgtgttgatacatcaactkvmcctgatattgagtggcc 590 . . . . 714bp
        |||||||||||||||||||||||| ||||||   |  |||||||| |||||
  1987 gtattgctgaaccgtgttgatacgtcaactgcacaggatattgaatggcc 1938
E. coli K12:
```

C.3 PCR PRIMERS DESIGN FOR DAIRY MARKER

TGATGAGTCCTGAGTAA**CTA**

5′ TGATGAGTCCTGAGTAA**CTA**TATTGTGAGGSYTNGCATAATGGCATTCAGAATGAGTG

ACAACCMCGGRNCCATAAAAATTTATAATYNNTGCTGGCCGGRANCTAATGAATTTATTG

GTGAAGGTGACGCATATATTCCGCCTCATACCGGTCTGCCTGCAAACAGTACCGATATTG

CACCGCCAGATATTCCGGCTGGCTTTGTGGCTGTTTTCAACAGTGATGAGGCATCGTGGC

ATCTCGTTGAAGACCATCGGGGTAAAACCGTCTATGACGTGGCTTCCGGCGACGCGTTAT

TTATTTCTGAACTCGGTCCGTTACCGGAAAATTTTACCTGGTTATCGCCGGGAGGGGAAT

ATCAGAAGTGGAACGGCACAGCCTGGGTGAAGGATACGGAAGCAGAAARRCTGTTCCGGA

TCCGGGAGGCGGAAGAAACAAAAAAAAGCCTGATGCAGGTAGCCAGTGAGCATATTGCGN

CCGCTTCAGGATGCTGCAGATCTGGAAATTGCAACGGAGGAAGAAACCTCGTTGCTGGAA

RCCTGGAAGAAGTATCGGGTATTGCTGAACCGTGTTGATACATCAACTKVMCCTGATATT

GAGTGGCCGACTTCACCTGCAGAGTAATCAACGGCGCANGGGATGATANTTTTGTGCTTK

RARKGNCCTCCCCGACCATTCAGATAGTCGGGGAGGGGGNNNTWTNGGGATTATNACGCA

ATCNNRGTTT**TAG**TTACTCAGACTCATCA3′

ATCAATGAGTCCTGAGTAGT

Key

*Mse*I+CTA primer (where CTA is the selective nucleotide overhang).

Synthetic Adapter Ligated to Restriction Fragment

Target-specific PCR primers: Dairy 1B

Dairy 2B

C.4 *PST* I DIGESTION OF THE 462 BP DAIRY FAECAL MARKER

(**Key**: PCR primers Dairy 1B and Dairy 2B)

Dairy 1B ⟶

1 to 75 bp

ATTGCACCGCCAGATATTCCGGCTGGCTTTGTGGCTGTTTTCAACAGTGATGAGGCATCGTGGCATCTCGTTGAA
TAACGTGGCGGTCTATAAGGCCGACCGAAACACCGACAAAAGTTGTCACTACTCCGTAGCACCGTAGAGCAACTT

76 to 150 bp

GACCATCGGGGTAAAACCGTCTATGACGTGGCTTCCGGCGACGCGTTATTTATTTCTGAACTCGGTCCGTTACCG
CTGGTAGCCCCATTTTGGCAGATACTGCACCGAAGGCCGCTGCGCAATAAATAAAGACTTGAGCCAGGCAATGGC

151 to 225 bp

GAAAATTTTACCTGGTTATCGCCGGGAGGGGAATATCAGAAGTGGAACGGCACAGCCTGGGTGAAGGATACGGAA
CTTTTAAAATGGACCAATAGCGGCCCTCCCCTTATAGTCTTCACCTTGCCGTGTCGGACCCACTTCCTATGCCTT

226 to 300 bp

GCAGAAARRCTGTTCCGGATCCGGGAGGCGGAAGAAACAAAAAAAAGCCTGATGCAGGTAGCCAGTGAGCATATT
CGTCTTTYYGACAAGGCCTAGGCCCTCCGCCTTCTTTGTTTTTTTTCGGACTACGTCCATCGGTCACTCGTATAA

301 to 375 bp *PstI*

GCGNCCGCTTCAGGATGCTGCAGATCTGGAAATTGCAACGGAGGAAGAAACCTCGTTGCTGGAARCCTGGAAGAA
CGCNGGCGAAGTCCTACGACGTCTAGACCTTTAACGTTGCCTCCTTCTTTGGAGCAACGACCTTYGGACCTTCTT

376 TO 450 bp *PstI*

GTATCGGGTATTGCTGAACCGTGTTGATACATCAACTKVMCCTGATATTGAGTGGCCGACTTCACCTGCAGAGTA
CATAGCCCATAACGACTTGGCACAACTATGTAGTTGAMBKGGACTATAACTCACCGGCTGAAGTGGACGTCTCAT

451 to 462 bp

ATCAACGGCGCA
TAGTTGCCGCGT

⟵ Dairy 2B

Note:

Theoretical digestion with *Pst* I indicates that three fragments should be obtained: 17, 123 and 322 base pairs. The 17 bp product would be too small to visualise on agarose.