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.