CHAPTER 4

DISCUSSION

4.1 General

Shigellosis or bacillary dysentery is mainly caused by *Shigella* spp. Determination of the etiology of this infectious disease has always been through microscopic examination, culture method and serologic identification. The organisms affect the host's gastrointestinal system and give rise to watery and/or bloody diarrhoea. Improper collection of the specimens and late processing of the faeces can thwart the success of attempts to isolate *Shigella* from faecal specimens. If the samples are not appropriately and rapidly delivered to the laboratory, *Shigella* spp. may not survive the pH changes in the faeces. Moreover, there is no reliably effective enrichment medium for all species of this fastidious organism. Even when *Shigella* is successfully isolated, the results of culture are available only after two days, whereas medical treatment must be decided upon when the patient is first seen. If samples are collected after antibiotic therapy, the growth of the cultured organism maybe impaired as a result of sublethal injuries by the antibiotics. Confirmation of the causative agent needs serotyping; however, diagnostic antisera are not easily available in most developing countries.

On the other hand, sensitive and specific PCR assays have been developed to amplify and detect DNA from *Shigella* spp. (Frankel *et al.*, 1990; Houng *et al.*, 1997; Vargas *et al.*, 1999). A primary advantage of this diagnostic method is the rapid turnaround time, which typically amounts to a few hours. The starting material for PCR can be obtained from any tissue or body fluid. Degraded DNA is also suitable as long as the target sequence is intact and it need not be in an amount equivalent to a viable inoculum. This powerful technique also offers several advantages over other DNA-related methods: PCR can amplify specific DNA sequences from as few as 25 bp up to 10 kb by using primer-specified target sequences instead of the entire genome. It is

several orders of magnitude more sensitive than direct hybridization and requires only a single DNA molecule that need not be highly purified (Atlas & Bej, 1994).

With many apparent advantages, the potential of PCR in the detection of infectious agents has been quickly exploited. Nevertheless, PCR assays have limitations. They require a validated protocol, sophisticated technology, training and rigorous quality assurance. Through improper handling of reagents and samples, contamination with PCR amplicons or cellular DNA can lead to false-positives. Lievano *et al.* (2002) used PCR to identify *Bordetella pertussis*, which causes pertussis or whooping cough. When comparing results from parallel tests from a private laboratory and the New York State Department of Health reference laboratory (NYSDOH/WC), there was a difference in PCR-positive cases reported. Of the 33 primary respiratory specimens that were tested by PCR for the presence of *B. pertussis* DNA by both laboratories, the private laboratory found 20 positive cases compared with 8 at NYSDOH/WC. The authors postulated that DNA contamination might have been exacerbated by the larger-than-normal amount of samples processed during an outbreak, and cautioned that PCR when well performed, is used as a valuable adjunct to culture and a presumptive test for *B. pertussis* detection. PCR cannot differentiate between dead and viable organisms and it is highly sensitive, therefore it is possible that viable and nonviable shigellae are detected in clinical specimens in the absence of symptomatic infection.

4.2 Comparison of DNA extraction methods

The rejection of the null hypothesis ($\alpha = 0.05$, $p = 0.00$) indicated there was a significant difference between the results obtained when boiling and phenol-chloroform extraction methods were used to prepare the bacterial DNA templates. In this study, DNA templates from the cost-efficient but less laborious boiling method gave higher

percentages of virulence-associated genes amplications than phenol-chloroform extraction.

Boiling easily disrupts Gram-negative bacteria such as shigellae and salmonellae, thus releasing the DNA for PCR amplifications. Since PCR does not require highly purified templates, crude DNA preparations like boiling, are adequate (Ivinson & Taylor, 1991; Kocher & Wilson, 1991). It does not need laborious preparations, hence saving time and costly reagents. There is no need for highly trained personnel or complicated equipment to prepare DNA template and the PCR results generated are equivalent to other DNA extraction methods.

This study did not employ a separate plasmid isolation protocol as disruption of bacterial cell walls via boiling was adequate to release both chromosomal and plasmid DNAs into the lysate. The successful detection of plasmid-encoded genes (i.e. *ial*, *ipaH* and *sen*) released by boiling, was as evidently reportedly in Chapter 3. An additional plasmid extraction process by way of conventional alkaline lysis method or using commercial kits would entail higher costs and a longer time to achieve results.

Numerous published reports on the use of PCR to detect bacteria had not found any detrimental effect of the boiling method. Hendolin *et al.* (1997) used it to prepare DNA template for the simultaneous detection of four bacterial species in middle ear effusions. They even advocated that the frozen boiled lysates could be used repeatedly, although for not more than ten times as there would be a weakening of PCR signals. Houng *et al.* (1997) boiled bacterial cultures and stool suspensions to detect and differentiate *Shigella* and EIEC strains. Boiled cell lysates were also successfully applied in the development of a seminested PCR method for the detection of *Shigella* in spiked environmental water samples (Theron *et al.*, 2001).

Phenol-chloroform extraction is a conventional technique that uses organic solvents to extract contaminants from cells. The cells are lysed using a detergent and

then mixed with phenol, chloroform and isoamyl alcohol. Contaminants are separated into the organic phase and DNA is extracted from the aqueous phase by alcohol precipitation. This is a time-consuming and cumbersome method. It uses toxic compounds such as phenol, and the isolated DNA may still contain residual phenol and/or chloroform, which can inhibit enzyme reactions in downstream applications such as PCR. Several compounds commonly used in this method were found to have some inhibitory effect on PCR (Rossen *et al.*, 1992). Phenol, SDS, lysozyme, NaOH, ethanol and EDTA hindered the efficiency of PCR at 0.5%, 0.01%, 0.5 mg/ml, 8 mM, 5% and 1mM respectively. Toxic waste that is generated must be disposed of with care and in accordance with hazardous waste guidelines.

Its unsuitability for high-throughput applications resulted in studies to explore other quicker and efficient methods of isolating DNA. It was found to cause more frequent cross-contamination than the GeneReleaser and QIAamp kits in the extraction of hepatitis B virus DNA from serum (Kramvis *et al.*, 1996). Wilde *et al.* (1990) reported that phenol-chloroform extraction failed to remove non-biological inhibitors in stool suspensions unless the specimens were further purified in chromatography medium CF11. Although Stacy-Phipps *et al.* (1995) found that DNA extracts obtained from the phenol-chloroform method required further dilution (at least a 1:9 000 dilution) to reduce the level of PCR inhibitors, Toye *et al.* (1998) cautioned that the usage of dilution to eliminate inhibitors could result in missing a PCR-positive specimen containing low numbers of target DNA copies.

4.3 Optimization processes

4.3.1 Monoplex PCR

An optimization process was carried out on monoplex PCR prior to combining all the primer sets in a multiplex PCR in order to:

i. verify the band size of the amplicon when each of the gene was primed by its primers and amplified accordingly,

ii. determine whether the resolving power of the recommended 2% agarose gel could distinctly separate the amplicon bands for *setIA* (309 bp) and *ial* (320 bp) when both primer sets were combined in a single reaction during mPCR. Should both bands are difficult to discriminate, re-designing of the primer sequences has to be carried out.

iii. obtain the optimum concentrations of PCR reagents and amplification conditions of each gene so as to provide a starting-point for the optimization work of the mPCR.

As stated in Section 3.2, initial runs of monoplex PCRs for the 5 primer sets were performed following the protocol by Vargas *et al.* (1999) who successfully amplified only the desired PCR products without primer-dimers or background noise. The failure of this study to reproduce the same result as Vargas *et al.* (1999) despite using the similar reagent concentrations and amplification conditions maybe attributed to the different brands of PCR reagents and primers used. The unspecific bands and bright primer-dimers obtained were possibly caused by a surplus concentration of primers used. The previous researchers had used *Taq* DNA polymerase from Gibco but did not specify the make of their primers, whereas all the PCR reagents in the present study were from Promega and primers from Integrated DNA Technologies. Different brands of enzymes display different specificity in PCR. Broude *et al.* (2001) showed *Taq* DNA polymerase (Amersham Pharmacia) produced more uniformly amplified products than AmpliTaq DNA polymerase (Applied Biosystems) but was accompanied by more primer-dimers. AmpliTaq also generated higher non-specific background than *Taq* DNA polymerase. After adjusting the concentrations of primers, dNTPs and $MgCl_2$

to 0.5 μM each, 130 μM each and 4 mM respectively, there was a uniform amplification of all genes without background noise (Fig 3.2).

4.3.2 Multiplex PCR (mPCR)

The quality of a multiplex amplification can be judged by the uniformity of amplification of different target sequences, the amount of primer-dimers and its clean background. The equal amplification of all genes in a single reaction ensures the absence of amplification suppression by one gene on another and lessens the possibility of having false-positives or false-negatives. The formation of primer-dimer artefacts and nonspecific products are substrates for PCR itself and they strive for enzyme, dNTPs and primers with the desired PCR products, causing reduced amplifications for the latter.

Specific, sensitive and consistent amplification of genes in a mPCR are dependent on factors such as primer concentrations, annealing temperature, the amount of enzyme, dNTPs and $MgCl_2$. There was uneven amplification of genes when all five primer sets were combined at a similar concentration each (0.5 μM). Both *set1B* and *ipaH* were strongly amplified but there was no *sen* amplification. Preferential amplification of one target sequence over another is a known phenomenon in mPCRs. Mutter & Boyton (1995) wrote that PCR selection is a mechanism that favours the amplification of certain targets due to the properties of the target, the target's flanking sequences or the entire target genome. These properties include inter-region differences in GC content that lead to biased accessibility of targets within genomes due to secondary structures and the gene copy number within a genome.

When the primer concentrations for both *set1B* and *ipaH* were decreased to 0.3 μM each, *set1A* and *ial* fixed at 0.5 μM each and the "weak" locus of *sen* was increased to 0.6 μM each, a faint band of 798 bp was visible amidst nonspecific

background. However, it is postulated that the "strong" loci of *set 1B* and *ipaH* were still curbing the *sen* amplification. When primer concentrations for the "strong" loci were lowered to 0.1 μM simultaneously with the increase of primer concentrations for *sen* to 0.9μM, a strong band of 798 bp was produced, but the background noise grew more intense as well. The increase in background noise maybe due to the high primer concentrations of *sen*, which had promoted mis-priming and resulted in the accumulation of non-specific products. Another evidence of high primer concentrations was the presence of bright primer-dimers at this juncture of study.

Henegariu *et al.* (1997) suggested that by raising the buffer concentration (indirectly the KCl concentration as well) from 1.4X to 2X at a maintained $MgCl_2$ concentration, the presence of long unspecific products could be decreased in a mPCR. Primer pairs producing longer PCR products performed better at lower buffer concentrations compared to those producing shorter products that gave better signals at higher buffer concentrations. Longer products became harder to denature at higher buffer concentrations, hence the preferential amplification of shorter molecules. This probably explained why at 1.4X buffer concentration, all the five desired fragments were seen with some unspecific bands but at 2.4X, only the shortest fragment (147 bp) was visible. Evidently, a higher buffer concentration did not improve the overall amplification of the genes in this study, as there was a wide range of fragment sizes involved from the shortest at 147 bp (*set1B*) to the longest at 798 bp (*sen*).

Similar to monoplex PCRs, mPCRs should be performed at a stringent temperature, allowing amplification of all genes of interest without background by-products. Previous experiments had shown that by lowering the annealing temperature by 4 to 6°C, loci could be co-amplified in multiplex mixtures (Henegariu *et al.*, 1997). The optimization of buffer concentration demonstrated that concentrations of 1.6X and 1.8X produced strong amplification for all five genes in the presence of nonspecific

bands. Thus, optimization of annealing temperatures from 49 to 59°C was carried out separately in the mentioned buffer concentrations. The higher buffer concentration had a "cleaner" background than the lower buffer concentration when annealing temperatures were increased in intervals of 2°C. Complexity and competitive nature of mPCRs over monoplexes were portrayed when an increase in the annealing temperature did not enhance the specificity of amplification reaction as seen at 1.8X buffer concentration(gel R). In fact, the intensity of an unspecific signal at ~ 700 bp was the greatest at the highest temperature. The temperature of 49°C was not chosen as the optimized annealing temperature (1.8X buffer concentration) because the *sen* amplification was comparatively weaker than the other genes and thus, maybe easily missed or may not be amplified during the screening process of bacterial strains or clinical specimens, giving rise to false-negative results. The annealing temperature of 55°C was chosen for further optimization processes with other untested PCR parameters because its 798-bp band was as intense as the ones amplified at 57°C and 59°C and the unspecific band was dimmer.

An unequal amplification of various loci in the presence of spurious bands can be caused by a high concentration of enzyme in PCRs. Although Henegariu *et al.* (1997) found that the most efficient enzyme concentration (Perkin-Elmer) was approximately 2 U in a 25-µl mPCR reaction volume and Chamberlain *et al.* (1990) worked with 5 U/50 µl (AmpliTaq) to co-amplify nine genes, there was a greater and more intense background noise when the enzyme concentration was increased from 2 to 4 U (Promega) in this study. Following the recommendation of Innis & Gelfand (1990) to test enzyme concentrations from 0.5 to 5 U/100 µl (= 0.125 to 1.25 U/25 µl), the amplification of 798-bp band was irregular in lower *Taq* concentrations of 0.6 to 1 U/25 µl reaction mixture. The enzyme concentration was difficult to optimize based on the recommendations of literature as *Taq* from different suppliers behaved

differently because of their diverse formulations. Its concentration had to be determined empirically during optimization of each PCR assay.

The availability of sufficient dNTPs ensures *Taq* fidelity. Excess dNTPs will reduce the amount of free Mg^{2+}, which will interfere with the enzyme activity and decrease primer annealing. Inadequate concentrations of dNTPs will affect the yield of the amplified products (Rolfs *et al.*, 1992). Even though the rule of thumb suggests that dNTP concentrations of 200 μM each be used, there was no difference in the results of this study when dNTPs were varied from 130 to 220 μM each. *sen* gene remained to have unsatisfactory amplification.

Despite exhaustive attempts to optimize the simultaneous amplifications of all five genes, only *set1B*, *set1A*, *ial* and *ipaH* had uniform and reproducible amplifications. These genes were unaffected by the optimizations of most PCR parameters. Although initial co-amplifications produced bright primer-dimers and unequal signal intensities amongst the four genes, a reduction of primer concentrations for the "strong" loci corrected the faults. The number of PCR cycles remained unchanged because after 30 cycles, most primers would have been converted into PCR products and reaction conditions appeared to favour the annealing of the 3'-OH ends of the products to DNA template or to themselves. The 3'-OH ends of the products would be extended to higher molecular weight DNA and were randomly terminated during the additional cycles causing smears on agarose gels (Bell & DeMarini, 1991). The high $MgCl_2$ concentration of 4mM optimized in monoplex PCR was retained in mPCR. Its concentration needed to be proportional to the amount of dNTPs and these values could be constant for any reaction (Henegariu *et al.*, 1997). Saiki (1989) also stated that any $MgCl_2$ concentration between 1 and 4 mM was usually adequate in an amplification process.

Adapting multiple amplifications in one reaction has always been difficult to achieve due to incompatibility of primers. It may then be necessary to subgroup the primers in smaller multiplexes or to re-design the primer sequences. The second option involved expertise or experience in primer designing, a waiting period for the new primers and optimizations of the new system. For these reasons, researchers involved in the development of mPCRs often work only with the genes that can be co-amplified satisfactorily and disregard the "uncooperative" gene(s), divide the primers into smaller mPCRs or abandon the whole project.

Bej *et al.* (1990) could not achieve equal amplifications of multiple genes in *Legionella* spp. The mPCR amplification of two different *Legionella* genes, *mip* and 5S rRNA, was achieved by staggered amplification: only the *mip* primers were used for amplification in the first seven PCR cycles, followed by addition of the 5S rRNA primers for the remaining 38 cycles. The addition of a sixth primer pair led to poor amplification of some products in the development of a mPCR of bacterial virulence factor genes (Chizhikov *et al.*, 2001). When relaxed PCR conditions such as a lower annealing temperature were attempted, there were amplifications of all genes together with additional non-targeted genes. Even when the conditions were more stringent, some of the multiplex bands were difficult to interpret. While attempting to co-amplify 11 virulence genes to detect *E. coli* infections, Pass *et al.* (2000) also reported that some primers interfered with the amplifications of other genes with the production of spurious bands or the loss of amplification for certain genes. Consequently, adequate amplifications were achieved in four multiplex systems.

4.4 Sequencing

The PCR products were sequenced to confirm that they are of the desired gene sequences and not that of some artefacts with similar molecular weights as *set1B*, *set1A*,

ial, *ipaH* and *sen*. Sequencing also provided an alternative method to discount any false-positives in the PCR results.

Although minor sequence diversities between the amplicons and published *set1B* and *ial* gene sequences from the GenBank were observed, the amplicons showed high percentages of sequence similarity (more than 95%) with the referred genes. Sequence divergence does occur in the same bacterium with related strains (Wang *et al.*, 2000) or among different strains (Kong *et al.*, 2002). Despite the fact that the PCR products in this study were amplified from the same species as the published sequences, *S. flexneri*, the particular isolate used in this study is not the same isolate as those registered in the GenBank. *set1B*, *set1A* and *sen* were sequenced from *S. flexneri* isolate 2457T, *ial* from *S. flexneri* 2a isolate YSH6000 and *ipaH* from *S. flexneri* strain 5. Hence, it is not surprising that the sequence similarities of *set1B* and *ial* in this study were not 100% the same as the published sequences.

Forbes *et al.* (1995) noted that gene variability that is seen in PCR analyses of individual genes maybe due to mutations that result in no change in a particular amino acid or mutations that may cause substitution of an amino acid that does not affect the function of the protein. These allelic variations could have arisen by two mechanisms: a progressive accumulation of point mutations in different strains over time or by horizontal gene transfer of sequences between strains, which allows numerous re-assortments of a few variant sequences.

The genomic diversity at gene level has been put to use in epidemiological studies. By means of sequence variations in the *uidA* (β-D-glucuronidase gene), Farnleitner *et al.* (2000) detected and differentiated populations of *E. coli* from environmental freshwater. A PCR-based denaturing-gradient gel electrophoresis (DGGE), which detects single base substitutions in DNA, and PCR with direct sequencing, correlated perfectly in the screening of 50 *E. coli* isolates. They found nine

polymorphic sites and an average 2.3% of pairwise differences between alleles of the *uidA* gene fragments. Sequences from 20 *E. coli* H7 *fliC* genes belonging to ten different serotypes were analyzed and Wang *et al.* (2000) found ten different sequences with pairwise differences from 0.06 % to 3.12%. By comparing the H7 *fliC* sequences with sequences of *fliC* from other *E. coli* H-antigen type strains, they were able to identify H7-specific primers, which enabled the development of a PCR-based method to identify O157:H7 strains rapidly and sensitively.

4.5 Reproducibility

The development of an amplification assay that is aimed to detect infectious agents must ensure the reproducibility of the method. All the 28 *Shigella* strains had reproducible results in the mPCR system. However, the single amplification of *sen* failed to have 100% reproducibility. The chances of some of its unrepeatable results were false-positives in the first round of amplification were remote because in all the amplifications, the negative controls were void of the *sen* presence. A likely explanation would be the amplified segment of the gene is located in an unstable region of the *inv* plasmid, causing it to be easily deleted during *in vitro* storage in laboratory conditions (Venkatesan *et al.*, 1989) and thus, its low-copy number could not be detected during the repeated PCR assays.

4.6 Specificity

The oligonucleotide primer pairs used in this study, ShET1A, ShET1B, ial, Shig 1 and Shig 2, and ShET2 had similar sequences with those previously reported (Table 2.2). Under the monoplex PCR and mPCR conditions described in Sections 2.2.3.1 and 2.2.3.2, only the positive control strain of *S. flexneri* 2a (TH13/00) generated PCR products with the expected molecular weights of 309, 147, 320, 423 and 798 bp

respectively using the above-mentioned primers. No amplification was observed from the other 12 non-*Shigella* strains. Nevertheless, it cannot be concluded that both the PCR assays were specific for *Shigella* strains only because studies by Venkatesan *et al.* (1989), Frankel *et al.* (1990) and Nataro *et al.* (1995) had established that the *ipaH*, *ial* and *sen* genes respectively, were also located in EIEC strains. Both *set1A* and *set1B* genes encoding for the ShET1 enterotoxin had not been found in EIEC strains yet (Nataro *et al.*, 1995).

4.7 Sensitivity

A comparison of mPCR sensitivity using bacterial cultures could only be carried out with other published reports on monoplex PCRs as there was a dearth of mPCR publications on *Shigella* spp. The detection sensitivity for the present study's mPCR, 2.45×10^5 cfu/ml of lysate, was 10-fold more sensitive than the PCR assay of Yavzori *et al.* (1994) who amplified with *ial* primers only. Theron *et al.* (2001) increased their PCR sensitivity limit from an initial 1.6×10^6 cfu/ml to 1.6×10^5 cfu/ml when the bacterial strains were further subjected to a seminested PCR assay with *ipaH* primers. Although Houng *et al.* (1997) developed a mPCR protocol to detect and differentiate *Shigella* infections, they had only reported the sensitivity limit for the detection of IS630-specific sequences of *S. sonnei*, which was at 7.4×10^4 cfu/ml. As they had also stated that at least 10^4 to 10^5 cfu/ml were essential to yield a positive PCR product, the present study's sensitivity limit could be deemed more superior because it detected the presence of four PCR products simultaneously at a minimum concentration of 2.45×10^5 cfu of shigellae /ml. Lampel *et al.* (2000) reported that amplification of the *ipaH* gene was observed from as few as 40 cfu in their study. The higher detection sensitivity maybe due to the incubation of the stock culture for 4 h to obtain a cell count

of 10^9 to 10^{10} cfu/ml compared to this study's stock culture of 10^8 cfu/ml without incubation.

4.8 Faecal-spiking and sensitivity

Processing of specimens is the most critical step in PCR assays. It is necessary to remove or inactivate inhibitory substances to PCR even though the isolated DNA need not be highly purified. Faecal specimens are amongst the most complex specimens for direct PCR testing as they contain inherent PCR inhibitors that are often co-extracted with bacterial DNA. Potential inhibitors present include heme, bile salts (sodium glycocholate and sodium taurocholate), bilirubins, which inhibit PCR at 50 µg/ml and complex polysaccharides (Monteiro *et al.*, 1997; Ruth, 1999). Moreover, the non-uniformity of faecal specimens in terms of matter, target organisms and associated background faecal flora can cause the yield and purity of DNA extracted from the specimens highly variable from specimen to specimen (Holland *et al.*, 2000).

Many methods for processing faecal specimens for PCR have been described. Suspending faecal samples in solutions followed by pelleting larger faecal debris by low-speed centrifugation is commonly performed. The clarified supernatant is then subjected to another centrifugation step and the resulting pellet is extracted for its DNA by proteinase digestion and alcohol precipitation (Frankel *et al.*, 1990).

Other long, complex and at times, costly methods with not easily available materials include capturing bacteria from faeces by using specific monoclonal antibody-coated superparamagnetic beads (Islam & Lindberg, 1992), the use of the chaotropic DNA glass matrix (Caeiro *et al.*, 1999) and commercial kits (Monteiro *et al.*, 1997). These techniques may give rapid and sensitive results but overall expenses and time are increased due to the many intricate steps or resources demanded. Certain chemicals

such as phenol and chloroform can inhibit PCR and the use of toxic or hazardous reagents poses a disposal problem.

In order to overcome the inhibitory effects of certain biological agents in faeces, this study employed an easy faecal culture enrichment step prior to DNA extraction by boiling. Although dilution of the faecal suspension before the enrichment step helped to lessen the inhibitors and at the same time, would make the mPCR less sensitive as fewer bacteria would be present in the diluted suspension, the short 4-h enrichment step would increase the total number of target sequences caused by more bacterial growth, and the overall detection sensitivity of the assay. Although PCR cannot differentiate between dead and viable bacteria, enrichment helped to dilute the concentration of dead bacteria, thus reducing the probability of detecting them by the subsequent mPCR assay.

The detection sensitivity limits of mPCR by using bacterial culture without preincubation and spiked faeces with preincubation were 2.45×10^5 cfu/ml and 5.0×10^4 cfu/ml respectively. Obviously, preincubation in BHI had helped to increase the sensitivity limit by 10-fold in the spiked faeces despite it having biological PCR inhibitors. Houng *et al.* (1997) also found preincubation increased the sensitivity of PCR. An inoculated faecal sample with 10^3 cfu of shigellae /ml was incubated for up to 8 h at 37°C. The faecal suspension was then subjected to PCR at different time points. The sample yielded a negative PCR result at time zero but a positive result after 4 h of enrichment. Kongmuang *et al.* (1994) and Hu *et al.* (1999) too attested the fact that an enrichment step lowered the concentration of PCR inhibitors in faeces but increased the number of organisms concurrently. The protocol represents a compromise between a decrease of PCR inhibitors in faeces as a result of dilution, and an increase of viable bacteria from enrichment.

A critical factor in combining cultivation in an enrichment broth with PCR is to select a medium that is not inhibitory to PCR. Stone *et al.* (1994) compared the

suitability of four enrichment media in the detection of *Salmonella* serovars from clinical samples by PCR. A 2-h incubation period in BHI and selenite-cystine broths increased the detection sensitivity to 80 cfu and 100 cfu of *Salmonella* in seeded faeces respectively. However, Rappaport-Vassiliadis and tetrathionate broths were inhibitory to PCR. BHI is not inhibitory to the growth of *Shigella* as it was routinely used to grow *S. flexneri*-derived strains in the work of Tamano *et al.* (2002). Yavzori *et al.* (1994) found that BHI was a more effective enrichment medium for shigellae than gram-negative (GN) broth. The number of PCR positive results using BHI was higher than GN: 19 positive results with *ial* primers in BHI compared to 4 positive results in GN, and with *virF* primer, there were 21 positive results in BHI incubation compared to 11 positive results in GN incubation.

A comparison of detection sensitivity limits achieved by spiked faecal sample in the present study and other published works could not be done objectively because of different kinds of PCR assays used, non-similar DNA templates with most studies focusing on environmental and food samples, the presence or absence of an enrichment step, type of medium used and length of incubation time.

There is a paucity of information on the application of mPCR in *Shigella*-spiked faecal specimens. Hence, the average sensitivity limit achieved by this study, 5.0×10^4 cfu of *S. flexneri* 2a /ml, could only be compared with the following reports on enrichment culture-monoplex PCR assays in faecal samples. The detection sensitivity limit attained by this study's mPCR system can be considered at par with those reports despite the simultaneous detection of four genes, as opposed to the usual single amplification of genes, from the same amount of template and reaction mixture. Following a 4-h incubation in BHI or GN broth, Yavzori *et al.* (1994) managed to detect up to $10^3 - 10^4$ cfu/ml by amplifying *virF* for the identification of *Shigella* in spiked faeces. Dutta *et al.* (2001) used *ipaH* primers to detect *Shigella* and EIEC strains in LB-

preincubated spiked faeces (4 h) at a sensitivity limit of 2 X 10^2 cfu/ml. Although there was no incubation, faecal specimens containing as a few as 10 cfu of *S. flexneri* yielded a positive hybridization result after 26 cycles of *ial* amplification (Frankel *et al.*, 1990).

Numerous experiments on spiked or seeded food and environmental water samples were carried out and nearly all reported an approximate sensitivity of 10^2 cfu/ml. In a mPCR developed for the simultaneous detection of *Salmonella* and *Shigella* using *invA* and *virA* sequences respectively in mussels, the detection limit was 10 – 10^2 cfu/ml of homogenate following a 22-h incubation in sterile buffered peptone water (Vantarakis *et al.*, 2000). Another mPCR assay for the identification of six bacterial pathogens in marine waters also reported a detection limit of 10^2 cfu for the amplifications of *aero*, *ipaH*, *ail*, *ipaB*, *epsM* and *Vpara* genes (Kong *et al.*, 2002).

Nested or semi-nested PCRs were often used to increase the sensitivity of PCR. During a single amplification by using *ial* primers, Lindqvist (1999) managed only to detect 0.5 – 1.0 X 10^5 cfu of *Shigella* /ml in seeded food compared to a 100-fold increase in sensitivity after a second round of amplification (1.0 X 10^3 cfu/ml). Theron *et al.* (2001) incubated artificially contaminated environmental water samples in GN broth for 6 h and in a semi-nested PCR, as low as 11 cfu of shigellae per ml of spiked tap water and 14 cfu/ml of spiked well water were detected by amplifying *ipaH* sequences.

Although it has been commonly reported that ingestion as low as 100 shigellae resulted in clinical disease, a study of the report by DuPont *et al.* (1989) found that the highest percentages of volunteers producing diarhoeal illness were at administered doses of at least 10^4 viable cells. Moreover, it had been reported that the numbers of *Shigella* in faeces of patients with clinical shigellosis ranged from 10^5 to 10^8 per gram of wet faeces (Dale & Mata, 1968). Thus, the average detection limit of mPCR described

in this study (5.0×10^4 cfu/ml) is within the common infectious dose of *Shigella* and less than the density shed by an infected individual.

4.9 Prevalence of virulence-associated genes in Malaysian strains

Precautions were taken to negate the possibilities of having false-positives or false-negatives in the amplification results. In order to ensure there was no false-negatives, pre-screening of a small selection of bacterial strains was carried out to find a suitable and consistent positive control strain (i.e. strain TH13/00). This particular strain was tested for its result reproducibility at least three times in different occasions. Optimizations of DNA extraction methodology and PCR conditions were also performed. If both of them were not fully optimized, true positive results might be missed during the screening process. At every screening, the positive control strain was freshly picked and boiled together with the other strains because the crudely extracted DNA would degrade more easily than its counterpart of high purity, even in storage at –20 °C. Therefore, should the positive control fail to give amplifications for all the five genes, it might be caused by partial DNA degradation, improper extraction procedures or some careless mistakes during PCR preparation.

At the same time, steps were also taken to rule out any false-positives during detection. A negative control consisting all the PCR reagents without bacterial template, was included in every amplification process. If there was contamination during the mixing of PCR reagents, it would be indicated in this "water-blank" negative control as well. Meticulous investigations would then be done to identify the source(s) of contamination before further PCR screenings were carried out. Using a new batch of PCR reagents sometimes helped to eradicate the problem too. Other procedures regularly followed to minimize wrong results included those reviewed in Section 1.7.3.

4.9.1 Overall prevalence of virulence-associated genes

Despite being proven that there was no significant difference between the prevalence of *set1A/set1B* with *ial* and *set1A/set1B* with *sen*, the prevalence of each gene in the Malaysian strains was in a decreasing order of *ipaH*, *set1A/set1B*, *ial* and *sen*.

The *ipaH* gene had been repeatedly shown to have a higher percentage of, if not a 100%, existence than the other genes in comparative studies to identify a more superior detection marker for shigellae (Venkatesan *et al.*, 1989; Lüscher & Altwegg, 1994). Its ubiquitous presence in all the Malaysian strains concurred with such studies. *ial* gene cluster is actually made up of four gene sequences called invasion plasmid antigen (ipa) genes (*ipaA*, *ipaB*, *ipaC* and *ipaD*) (Sethabutr *et al.*, 1993) and is present on the *inv* plasmid of virulent *Shigella* spp and EIEC strains (Venkatesan *et al.*, 1989). The *ipaH* gene, another antigen-coding gene, is also located on the similar plasmid but is not part of the *ial* gene cluster (Venkatesan *et al.*, 1989). Due to its presence on the *Shigella* chromosome as well, it maybe less compromised by spontaneous plasmid loss events.

sen gene is also found in all four species of *Shigella* that harbour the *inv* plasmid. Although Yavzori *et al.* (2002) recorded a 91% presence of this gene, Nataro *et al.* (1995) 83% and Vargas *et al.* (1999) 41%, the Malaysian strains had an even lower prevalence for it (~31%), which incidentally was the least detected amongst the five genes in this study. One explanation maybe that selective deletions of the *sen* location on the *inv* plasmid had taken place during storage/subculturing in the laboratory. The 100% correlation of both *set1A* and *set1B* detected in the same strains was another proof to the concept that ShET 1 enterotoxin consists of two subunits (subunit A encoded by *set1A* and subunit B encoded by *set1B*) separated by only 6 bp (Fasano *et al.*, 1995).

4.9.2 Distribution of virulence-associated genes according to species and serotype

The prevalence of *Shigella* species in this study shared some similarities to the global distribution (Kotloff *et al.*, 1999). It was reported that the majority of *Shigella* strains from developing countries are *S. flexneri*, which constituted for 76.4% of strains in this study, with *S. sonnei* being the next most common (13.6%). Although *S. dysenteriae* and *S. boydii* may occur equally frequently in other developing countries, 9.1% of the studied strains were from *S. dysenteriae* and 0.9% from *S. boydii*. Jegathesan (1984) also reported a similar order of prevalence of *Shigella* species in Malaysia from 1980 to 1981: 86.4% of strains were from *S. flexneri*, 13.3% from *S. sonnei* and 0.2% from *S. dysenteriae*. Our two neighbouring countries gave slightly different representations of prevalence of species from this study. The order of prevalence of species in 506 *Shigella* strains from Singapore corresponded to the global distribution reviewed by Kotloff *et al.* (1999). *S. flexneri* was the most common species (60.3%), followed by *S. sonnei* (33.6%), *S. boydii* (3.2%) and *S. dysenteriae* (3.0%) (Lim & Tay, 1991). Oyofo *et al.* (2002) recorded 80.9% of isolated *Shigella* strains were from *S. flexneri*, 14.6% from *S. sonnei* and 4.5% from *S. dysenteriae* in a study on the prevalence of faecal pathogens in Jakarta.

The distribution of serotypes in this study also differed from the global picture (Kotloff *et al.*, 1999). Of the 84 *S. flexneri* strains identified in this study, the predominant serotypes were 2a, 3a and 3c. However, the previous Malaysian report (Jegathesan, 1984) found that the distribution was in the descending order of serotypes 2a, 1b, 3a, 3c, 1a, 6, x and y. Kotloff *et al.* (2002) wrote that the most predominant *S. flexneri* serotype distribution from developing countries is serotype 2a, followed by serotypes 1b, 3a, 4a and 6. The Singapore study showed that the most isolated strains from *S. flexneri* were of types 2a and 1b (Lim & Tay, 1991). No serotypes were available for the Indonesian *Shigella* isolates.

Only *S. dysenteriae* 2 and *S. boydii* 6 were isolated in the present study although *S. boydii* types 1 and 5 and *S. dysenteriae* types 1 and 3 were predominant in Singapore (Lim & Tay, 1991). *S. dysenteriae* 1 is largely found in India, Nigeria and Singapore, while type 2 predominates in Guatemala, Hungary and Yemen. *S. boydii* 14 is most often isolated in India, Nigeria, Yemen, and type 2 in Guatemala (Kotloff *et al.*, 1999).

Both *set1A* and *set1B* were originally identified in a *S. flexneri* 2a strain (Fasano *et al.*, 1995). Follow-up work by Noriega *et al.* (1995), Talukder *et al.* (2002) and Yavzori *et al.* (2002) and the present study arrived at a similar finding: both genes were present almost exclusively in *S. flexneri* 2a. Both genes were detected in approximately 87% of *S. flexneri* 2a strains and had a bare presence in *S. flexneri* 3a and 4a of the Malaysian strains. None of the remaining strains yielded a positive result for them. The prevalence of these genes was significantly associated with *S. flexneri* 2a serotype based on the χ^2- test ran in this study ($p = 0.00$). Noriega *et al.* (1995) had also noted that the genes were minimally present in *S. flexneri* types 1a, 2b and 3b, and *S. boydii* in their study. None of their ten clinical isolates of EIEC hybridized with *set1* probe, suggesting that those strains did not elaborate the enterotoxin despite their almost homologus genetic background with *Shigella*. Yavzori *et al.* (2002) had no amplification of the *set1* gene in their EIEC isolates as well.

The ShET1 enterotoxin is made up of two distinct, yet contiguous, ORFs encoding two proteins of 7 and 20 kDa (Fasano *et al.*, 1995). The primers used in this study were designed to amplify either the *set1A*, encoding for the bigger protein or the *set1B*, encoding for the smaller protein (Vargas *et al.*, 1999). It was observed that all the strains registering a presence for *set1B* also yielded a positive result for *set1A*, which was supportive of the finding by Fasano *et al.* (1995). The latter had also discovered that ShET1 was chromosomally-encoded, as plasmid-cured derivatives of *S. flexneri* 2a strain also showed enterotoxic activity in ileal loops and Ussing chambers.

The invasion-associated locus (*ial*) is present in all *Shigella* spp. and EIEC strains that harbour the large *inv* plasmid. This was clearly demonstrated in the present study with its presence detected in various serotypes of *S. flexneri*, and *S. sonnei* and *S. dysenteriae* 2. The χ^2- test was unavailable to provide a significant relationship between the prevalence of *ial* with *S. flexneri* 2a serotype ($p = 1.00$) in spite of 42.2% of positive results were found in *S. flexneri* 2a alone. Not all the strains in this study showed the presence of *ial* as reported by Lüscher & Altwegg (1994) too. A loss of the *inv* plasmid may result in the absence of *ial* – this justification will be discussed at length in Section 4.9.3.

The observation of the absolute presence of *ipaH* in all the Malaysian strains, regardless of species and serotypes, was in agreement with reports by Sethabutr *et al.* (1993), Lüscher & Altwegg (1994) and Vargas *et al.* (1999). Venkatesan *et al.* (1989) had shown that this gene is available in multiple copies on both the chromosome and the *inv* plasmid of *Shigella* spp. and EIEC strains.

Another virulence-associated gene to be detected in members of all *Shigella* spp. and EIEC strains which have the large plasmid, is the *sen* gene. It was originally discovered in an EIEC isolate and had been localized to the large *inv* plasmid in 75% of EIEC strains and 83% of *Shigella* spp. (Nataro *et al.*, 1995). Based on the results of this study, the prevalence of *sen* gene was not significantly associated with *S. flexneri* 2a serotype ($p = 0.10$). It was also detected in *S. flexneri* 3a, 3c, 6 and *S. dysenteriae* 2. Interestingly, although Vargas *et al.* (1999) had found *sen* in ~58% of their *S. sonnei* isolates and Yavzori *et al.* (2002) in all their *S. sonnei* strains, no *sen* was detected in our *S. sonnei* strains. This maybe explained by an earlier observation that its plasmids were often lost at a frequency of ~50% per generation (Sansonetti *et al.*, 1982).

4.9.3 Distribution of virulence-associated genes according to year of isolation

One limitation of this study could be seen from the uneven distribution of *Shigella* strains isolated over the years. A huge discrepancy in the distribution arose from outbreaks of shigellosis in two states in 1998. Therefore, only comparisons of numerical values instead of proven statistical association between the prevalence of each virulence-associated gene and the year of isolation could be made in the following discussion.

More *Shigella* strains had positive amplifications for both the tandem genes (*set1A* and *set1B*) than for *sen* from 1997 to 1999, which maybe attributed to the stability of the respective gene location in the bacteria. The *set1A* and *set1B* genes had been located on the chromosome, which were more stable than any plasmid-encoded gene. They were less susceptible to *in vitro* losses due to storage/subculturing. In spite of a recent study that discovered that these genes were borne on a deletable chromosomal element designated as the *she* PAI (Rajakumar *et al.*, 1997), their rate of deletion at 10^{-5} to 10^{-6} was lesser than the plasmid's rate of total loss or deletion formation at 10^{-2} per bacterium per generation (Sansonetti *et al.*, 1982). The deletable positions of *set1A* and *set1B* could also justify for the remaining 12.8% of *S. flexneri* 2a in this study that showed absence for the genes, despite other reports of a 100% presence in all alike strains (Noriega *et al.*, 1995; Yavzori *et al.*, 2002).

An important feature of the virulence factors in shigellae is invasion and proliferation in colonic epithelial cells. This process is mediated in part by a 180- to 210-kb plasmid that has been termed the *inv*asion or *vir*ulence plasmid (Sasakawa *et al.*, 1986; Frankel *et al.*, 1990). The observations that both *ial* and *sen* are located on this highly unstable plasmid, prone to a total loss or selective deletion in subculture/storage at a rate of 10^{-2} per bacterium (Sansonetti *et al.*, 1982; Venkatesan *et al.*, 1989; Nataro *et al.*, 1995), are supported by the following findings. Table 3.8 showed that nine

strains recorded a presence for both *ial* and *sen* genes in 2000 as opposed to five strains in 1997. As the period between the year of isolation (1998 to 2000) and the year of the present study (2002) increased, so did the number of strains negative for both genes. Yavzori *et al.* (2002) also found that all their isolates that were stored for 12 years were negative for *sen* but were positive for *set1*.

Although Sasakawa *et al.* (1986) found that the *ial* gene cluster resides near a region of the plasmid that is a hot spot of spontaneous deletions, this study suggests that *sen* might be located on an even more unstable spot than *ial* on the plasmid. From 1997 to 2000, there were 21 strains that yielded a presence for *ial* but none for *sen* as compared to only ten strains, which had *sen* amplification, but none for *ial*. The total number of *sen* amplifications in this study (34/110 strains) was the least amongst the five virulence-associated genes, and this lends credibility to the above presumption.

On the other hand, *ipaH* is available in multi-copies on both the plasmid as well as the chromosome of *Shigella* (Venkatesan *et al.*, 1989). In agreement with this observation, there was 100% *ipaH* amplification irrespective of the year of isolation in this study as well as in the report by Vargas *et al.* (1999) who worked with *Shigella* strains isolated from 1993 to 1998.

4.9.4 Analysis of the profiles of virulence markers (pathotypes) among Malaysian *Shigella* spp.

The analysis revealed that a large proportion of the Malaysian *Shigella* strains (55/110) did not demonstrate the presence of either *ial* or *sen*, both of which are plasmid-encoded. Since invasiveness is a prerequisite for virulence in shigellae and most of these virulence-associated genes are located on the *inv* plasmid, these strains would have possessed the plasmid when first isolated from patients. Due to its spontaneous loss, which was a more likely reason as those strains were negative for

both *ial* and *sen*, or selective deletion of certain regions, these strains were rendered non-invasive and consequently avirulent by the time of this study. The instability of the plasmid was also evident from this particular analysis: approximately 31% of the strains presented only the *ipaH* gene, and 21/110 strains in the next most frequent profile had amplifications for the chromosomally-encoded *set1B/set1A/ipaH*.

The combination of *ial/ipaH* (14/110 strains), which recorded a higher presence than *ipaH/sen* (5/110 strains) lends additional weight to the suggestion that the locality of *ial* is probably less subjected to deletions than the location of *sen*.

ipaH was a part of all the five combinations of gene markers in this study with 34/110 strains yielding a presence for it singly. Its superiority in detection of shigellae by virtue of its multiple-copy presence was also obvious from the studies of Venkatesan *et al.* (1989), Sethabutr *et al.* (1993), Lüscher & Altwegg (1994) and Gaudio *et al.* (1997) who had all compared its efficiency with the *ial* gene cluster. However, Venkatesan *et al.* (1989) also pointed out that the hybridization of an *ipaH*-probe to a *S. sonnei* strain without its *inv* plasmid, had reflected that the gene was not an indicator of the invasive phenotype, but was specific and outstanding for the general detection of all *Shigella* spp. and EIEC.

4.10 Clinical specimens

Negative amplifications for the *Shigella* virulence-associated genes in the patients' faecal samples despite both the positive controls registering presence for the genes, confirmed the results of the culture-negative MacConkey and *SS* agar plates. The mPCR assay had complemented the microbiological method in determining that none of the patients had suffered from shigellosis.

4.11 Implications of this study

The present mPCR assay developed has several important implications:

i. It can detect virulence-associated genes located on the *Shigella* chromosome (ie. *set1A*, *set1B* and *ipaH*) and on the large *inv* plasmid (*ial* and *ipaH*) simultaneously, thus enabling the detection of *Shigella* spp. regardless of the location of the genes, in one reaction. This special feature is important in the detection of virulent and avirulent shigellae because the large plasmid is predisposed to loss or selective deletions during storage/subculturing.

ii. Asymptomatic carriers of *Shigella* who serve as potential reservoirs of the bacteria silently transmitting the disease within communities can be readily identified by the presence of *ipaH* gene. Shigellae may remain undetected in their faeces when examined by the conventional culture technique due to the low presence of organisms and certain non-culturable forms of the bacteria.

iii. The clinical applicability of the mPCR had also been shown in faecal-spiking tests. Its average detection sensitivity limit of 5.0×10^4 cfu of shigellae /ml was favourable to other studies. Furthermore, collection of bacterial cells from the enrichment broth by centrifugation followed by boiling to lyse the bacteria proved to be a simple, rapid and reliable method for template preparation. BHI was also cheaper than commercial extraction kits used in many related studies. The mPCR also did not involve any expensive chemical extraction or purification steps, which might increase the overall cost of the assay.

4.12 Limitations of the present study

This study was unable to concur or to disprove the findings by Venkatesan *et al.* (1989) and Nataro *et al.* (1995) that *ial*, *ipaH* and *sen* were specific for EIEC because such strains were not available for the specificity test at the time of study.

Both the amplification systems cannot distinguish *Shigella* species as *ial*, *ipaH* and *sen* could also be detected in all the four species. Hence, these virulence-associated genes are poor epidemiological markers to determine *Shigella* spp. Although *set1A* and *set1B* were primarily found in *S. flexneri* 2a, their presence had also been reported in *S. boydii* by Noriega *et al.* (1995).

The number of strains studied was made available by the IMR during the duration of research and represents a fraction of the total number of *Shigella* spp. received by the institute. The actual prevalence of the virulence-associated genes could be higher than what is determined in this study. In addition, the number of strains sent to IMR for serotyping is an underestimate. Many patients especially those from the rural areas do not have easy access to hospitals that are largely situated in big towns. Moreover, stoic Malaysians are likely to be indifferent to a few bouts of diarrhoea and will bear with the symptoms. Shigellosis is also self-limiting and usually resolves within a week in healthy individuals. The organism is fastidious in *in vitro* growth and successful isolation of shigellae is also hampered by the lack of skilled medical laboratory technicians or microbiologists in the health institutions. Statistical analyses could not be carried out to decide if there was an association between the prevalence of each gene with the year of isolation due to an unequal data distribution.

The efficacy of mPCR assay could only be ascertained in spiked faecal samples as there was no *Shigella*- positive faecal specimen from patients at the time of testing.