Subtyping of Shiga Toxin 2 Producing *E. coli* using Real-time PCR with HRMA

Master's Thesis by Solveig Hansen

Principal supervisor: Professor Jan Egil Afset
Co-supervisor: Professor Kåre Bergh

Norwegian University of Science and Technology
Faculty of Medicine
June 2015

Acknowledgements

I am deeply grateful for being given the oppertunity to do this project where I have learned so much. I give my sincerest thanks to the awesome Professor Jan Egil Afset who have been patient and supportive at all times. Without his help and advice I would never have been able to complete the project.

I would also like to thank Christina Gabrielsen for friendly advice and constructive critisism during all phases of the project.

Lastly, I thank Hilde Lysvand, Kjersti Haugum, and all the people at the Department of Medical Microbiology who helped me with valuable advice and to find my way in the lab.

Abstract

Background: Shiga toxin producing *E. coli* (STEC) can produce two types of Shiga toxins (Stxs): Stx1 and Stx2. The latter is more associated with severe disease such as haemolytic uremic syndrome (HUS) in humans, and is the main focus of this project. Stx2 can be categorised into several subtypes which differ in virulence: Stx2a, Stx2b, Stx2c, Stx2d, Stx2e, Stx2g, and Stx2f, with Stx2a and Stx2c being more virulent than other subtypes. Because of this it is desirable to be able to differentiate between them.

Aim: Two common methods used in Stx2 subtyping includes conventional PCR with agarose gel electrophoresis and sequencing. As these methods are laborious and relatively expensive, the objective of this project was to find a more effective way of subtyping Stx2 producing *E. coli*.

Materials and Methods: Real-time PCR with high resolution melt analysis was chosen as a method for Stx2 subtyping, and two different sets of primer pairs were tested and evaluated on seven reference strains provided by Flemming Scheutz and Statens Serum Institut in addition to a collection of pure and mixed culture samples containing Stx2 provided the Department of Medical Microbiology at St. Olavs Hospital.

Results: The first set of primer pairs could not successfully discriminate between the different Stx2 subtypes. The second set showed more promise as it could differentiate between subtypes Stx2a, Stx2c, Stx2d, Stx2b, Stx2e, Stx2g, and determine that a sample contained Stx2a and Stx2c or Stx2a and Stx2d when using the reference strains. Using the same set, 25 out of 29 pure culture samples were successfully subtyped as well as 18 out of 21 mixed culture samples.

Discussion and Conclusion: There were some problems regarding primer efficiency, and some of the tested patient samples were incorrectly subtyped or did not get amplified at all. Untill these issues are resolved, the method should not be put into practice.

Contents

1 Introduction

1.1 Shiga Toxins and *E. coli*

Shiga toxins (Stxs), also known as Shiga-like toxins (SLTs) and Vero cytotoxins (due to their effect on Vero cells) [1], are produced by bacterial species such as *Acinetobacter haemolyticus* [2], *Enterobacter cloacae* [3], *Citrobacter freundii* [4] as well as certain strains of *Escherichia coli* [5–7]. Shiga toxin producing *E. coli* (STEC) that have a harmful effect on humans (enterohemorrhagic *E. coli*, EHEC) can cause bloody diarrhea [8], hemorrhagic colitis [9], and haemolytic uremic syndrome (HUS) [10, 11].

STEC was first discovered in the USA in 1983 [12]. In Norway, it was first observed in 1992. There has been a steady incease in the number of reported EHEC cases in Norway from 26 in 2006 to 151 in 2014. In this time period there have been three deaths attributed to EHEC, all children that developed HUS. Many of the reported EHEC cases have been found coincidentally in healthy persons during contact tracing in relation to outbreak investigations. In 2006 there was an outbreak with 17 reported cases, 10 of which were children that developed HUS and one of whom died. In Europe, the most notable outbreak of EHEC occurred in Germany 2011 with 4397 reported cases and 51 deaths. [13]

The reservoar of STEC is the intestines of ruminants such as cattle and sheep [14, 15]. Infection of humans occurs mainly through contaminated water and poorly processed food, direct or indirect contact with animals, and between humans [13].

There are two antigenically distinct types of Stxs: Stx1 and Stx2 [16–18], each containing several subtypes. Stx1 subtypes include Stx1a [19], Stx1c [20], and Stx1d [21]. Stx2 subtypes include Stx2a [19], Stx2b [22], Stx2c [23], Stx2d [24], Stx2e [25], Stx2f [26, 27], and Stx2g [28]. In addition to this there are multiple variants within a subtype that differ only by a small number of amino acids. The Stx2 nomenclature has recently been revised and is described by Scheutz *at al* [29]. An example of a variant is Stx2c-O157-E32511, where Stx2 indicates type, c the subtype, O157 the serotype and -group, and E32511 the strain.

stx1 and *stx2*, coding for Stx1 and 2 respectively, are generally located on prophages called Stx-phages which are related to lambdoid bacteriophages [30, 31]. Stx-phages have double-stranded DNA genomes and a complex morphology of a head with icosahedral symmetry, a helical tail, and tail fibers [32].

Typical examples of Stx-phages are phages H19B and 933W, responsible for the *stx1* and *stx2* genes respectively. The relative locations of *stx2* and *stx1* in the bacteriophage genomes are illustrated in figure 1. The *stx* genes are inserted in the *E. coli* chromosome, commonly in genes *yehV* and *wrbA* respectively [33, 34]. Here they enter into a lysogenic replication cycle until external stimuli cause them to enter a lytic cycle [32].

Studies have shown that various types of antibiotics can cause an increased production of Stx. When antibiotics come in contact with STEC and the bacterium is threatened, the prophage can initiate an SOS response leading to activation of the lytic cycle and late genes where Stx is located [32, 35].

Interestingly, this relationship between bacterial host and bacteriophage have been suggested to maybe be beneficial to the bacterial population which the infected bacterium arised from. If the infected bacterium is threatened by competing bacteria, the virus may initiate the lytic cycle, releasing Stx-toxin which the competing bacteria may be sensitive to, effectively reducing the competing bacteria [32].

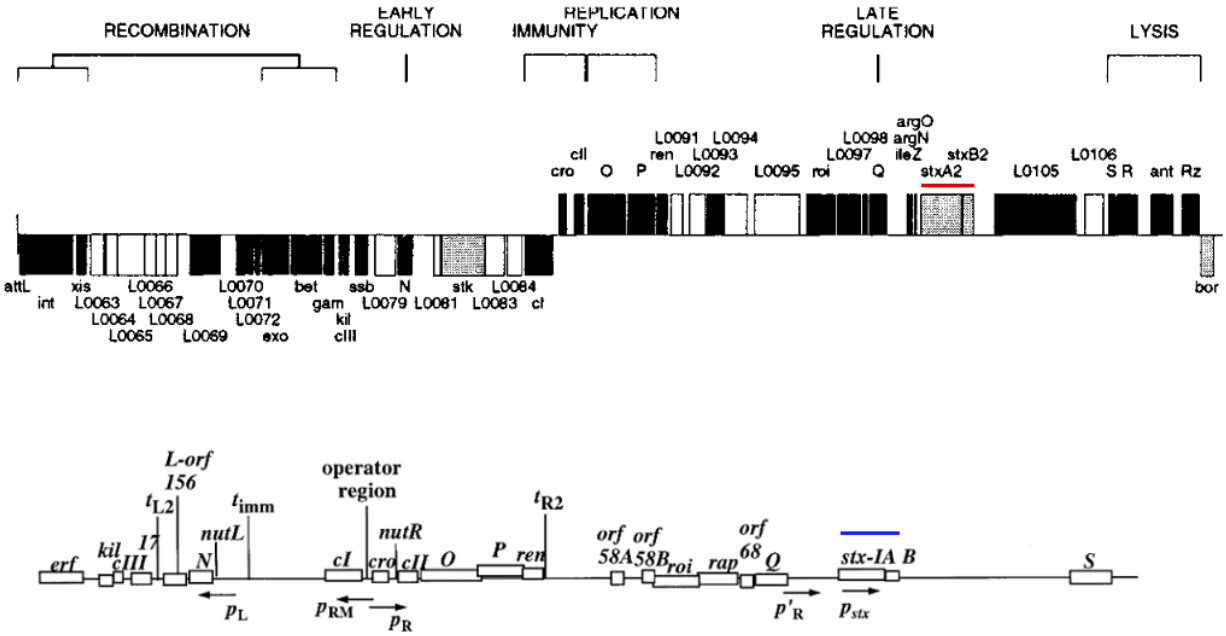


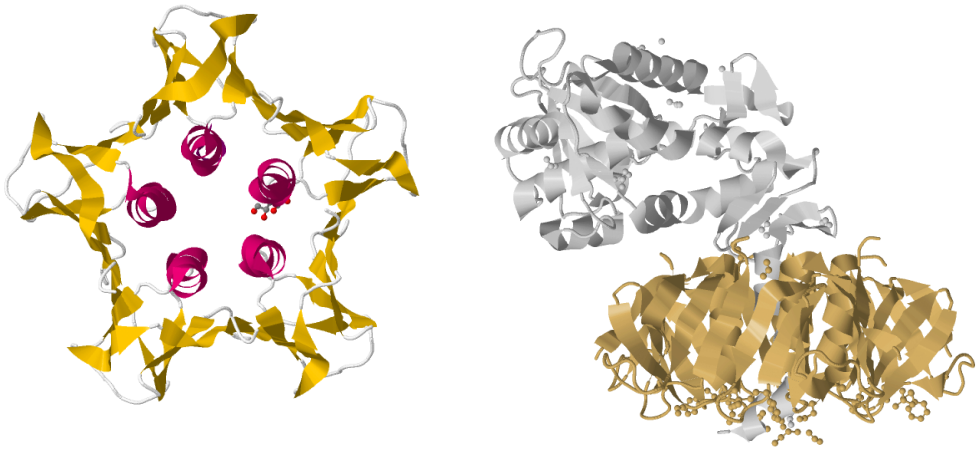
Figure 1: Top: Part of bacteriophage 933W genomic map showing the location of *stx2*, marked red [34], with permission from author. Bottom: Part of bacteriophage H-19B genomic map showing the location of *stx1*, marked blue [33], with permission from author.

Many STEC *stx* alleles are not associated with inducible prophages. These non-inducible Stx genes are thought to be part of cryptic prophages [36]. Some *E. coli* can be infected with multiple types of Stx-prophages [37], and it should be noted that a patient may harbour multiple Stx2-subtypes simultaneously [38].

Stx1 and Stx2 are AB_5-type proteins which are characterised by an enzymatically active A subunit (Stx2A) and a binding B domain (Stx2B) that is responsible for adhesion to host cells. *stx2A* is about 960 base pairs and code for Stx2A, and *stx2B* coding for Stx2B is about 260 base pairs. *stx2A* lies upstream of *stx2B*, separated by an intergenic sequence of about 10 base pairs. The B domain consist of five identical subunits forming a circle in which the C-terminal of the A domain is non-covalently anchored [16, 39], illustrated in figure 2. The B domain makes the bacteria able to adhere to host cells through binding to globotriacylceramide (Gb3), a glycosphingolipid [40]. The A subunit is responsible for removal of an adenosine residue (A-4324) from 28S rRNA, disabling eukaryotic large ribosomal subunit 60S, and consequently protein production [41].

Stx1 and Stx2 are structurally similar to Shiga toxins produced by *Shigella dysenteriae*. Stx1 is identical to the Shiga toxin produced by *S. dysenteriae* except for one residue in the A domain, while Stx2 shows approximately 56 % amino acid identity to Stx1 [44].

EHEC harbours a strecth of DNA or "pathogenicity island", called locus of enterocyte effacement, or LEE [45]. This is a collection of genes which help with bacterial adhesion to animal intestines, in addition to effacement [46]. Adhesins, such as intimin encoded by *eae* and located in LEE, play an important part when it comes to infection [46] and the *eae*-gene in combination with *stx2*-genes has been linked to the development of bloody diarrhea and HUS [35, 38]. It has also



(a) (b)

Figure 2: (a) A cartoon model of the Stx2 B domain forming a circle [42], with permission from author. (b) The B domain circle (brown) seen from the side with subunit A's C-terminal tail (grey) non-covalently anchored in the B domain [43], with permission from author.

been suggested that an interplay between Stxs in combination with intimin may help the bacteria attach to epithelial cells [47]. In the absence of intimin and LEE the bacteria must rely on other mechanisms for colonisation. An example is the use of aggregative adherence fimbrial pili encoded on an enteroaggregative *E. coli* plasmid [48].

Stx2 is more associated with the development of HUS than Stx1 [35, 38, 49]. One proposed reason for this is that the active site in Stx2 is more accessible [43]. Another is that Stx1, in contrast to Stx2, binds to toxin resistant cells in the lung rather than the more susceptible cells of the kidney [50]. Host factors such as young age is also a risk factor for development of serious disease [35].

Stx2a and Stx2c have been reported to be more closely related to the development of serious disease than other Stx2 subtypes [51–54]. Stx2b, Stx2d and Stx2e are associated with a milder course of disease but have still been found in some patients with HUS [55–58]. Stx2g and Stx2f are relatively rare and not assiciated with disease, and Stx2f is probably not a human pathogen [50, 53].

Some of the early means to detect Stx2 were based on antigenic differences between Stx1 and Stx2 [59] and focused on being able to separate them. As more Stx2-subtypes were uncovered, and showed differences regarding virulence, it became desirable to differentiate between these. Lately, conventional PCR [29] and sequencing is used in subtyping.

1.2 PCR

The polymerase chain reaction, PCR, is a method used to amplify a specific region of a DNA or RNA sequence, making millions of copies of this region. The reaction components consist of template DNA containing the region that is to be amplified, two short oligonucleotides, or primers, which can bind to the template DNA, a DNA polymerase, dNTPs, and a buffer mixture containing $MgCl_2$ among other components. These reaction components are mixed and then sequentially

subjected to three different temperatures referred to as denaturation, annealing, and extension, which together comprise a cycle. [60, 61]

During denaturation, the temperature is raised to circa 95 °C for 30-120 seconds to break the hydrogen bonds between the template DNA strands, making them single-stranded. The temperature is then lowered to 45-68 °C for about 15-30 seconds, allowing each of the two primers to hybridise to their respective strands, flanking the region of interest. By raising the temperature to about 70 °C for 15-30 seconds, the DNA polymerase is able to bind to the hybridised primers and extend their 3-ends, synthesising new double-stranded DNA, see figure 26. An extra preheating step can be included at the beginning of the PCR to ensure that all the DNA is properly denaturated, and an extra elongation step can be included at the end to make sure all the strands are completely synthesized. Theoretically there should be a doubling of PCR products after each cycle, but eventually amplification will come to a halt and a plateau is typically reached after 30-50 cycles, see figure 4. [61]

DNA molecules of different sizes can be separated and analysed by use of agarose gel electrophoresis. Agarose is a polysaccharide isolated from seaweed. Cooking agarose mixed with a buffer solution creates a liquid mixture that, when cooled, will form a solid, dense network of fibers with microscopic pores. Submerging the resulting gel in a buffer solution and applying an electric field will cause negatively charged DNA applied to wells in the gel to migrate from the negative electrode towards the positive. Larger DNA molecules will encounter more resistance travelling through the gel and travel a shorter distance than smaller ones. To be able to see how far the DNA has travelled, a fluorescent compond that binds to DNA and emits light when struck by UV-lighting can be used. By applying a solution containing DNA fragments of known sizes, a ladder, to the gel before the eletrophoresis, the size of the DNA fragment in question can be estimated by comparing its relative position to that of those in the ladder. [61]

1.3 Real-Time PCR

Real-time PCR is the continous analysis of PCR-product made during amplification [62], not an end-point analysis of the plateau phase as conventional PCR with gel electrophoresis, see figure 4. Real-time PCR is more sensitive than conventional PCR, allows for a more precise quantification of DNA, and there is an opportunity to study the melt profile of the PCR product. Real-time PCR does not provide information about the size of the PCR product, nor if there are multiple products present, but since it is a non-destructive method, agarose gel electrophoresis can be performed after real-time PCR to answer these questions.

A fluorescent dye can be used to quantify the relative amount of PCR product created during PCR amplification by using UV-illumination and a sensor that detects the emitted light. The dye will only emit light if it is bound to double stranded DNA. Consequently, a higher amount of double-stranded DNA will bind more dye, creating a stronger signal. This concept is used to visualise the gradual increase of double stranded product created in PCR by doing multiple readings of the fluorescence signal during amplification. [61]

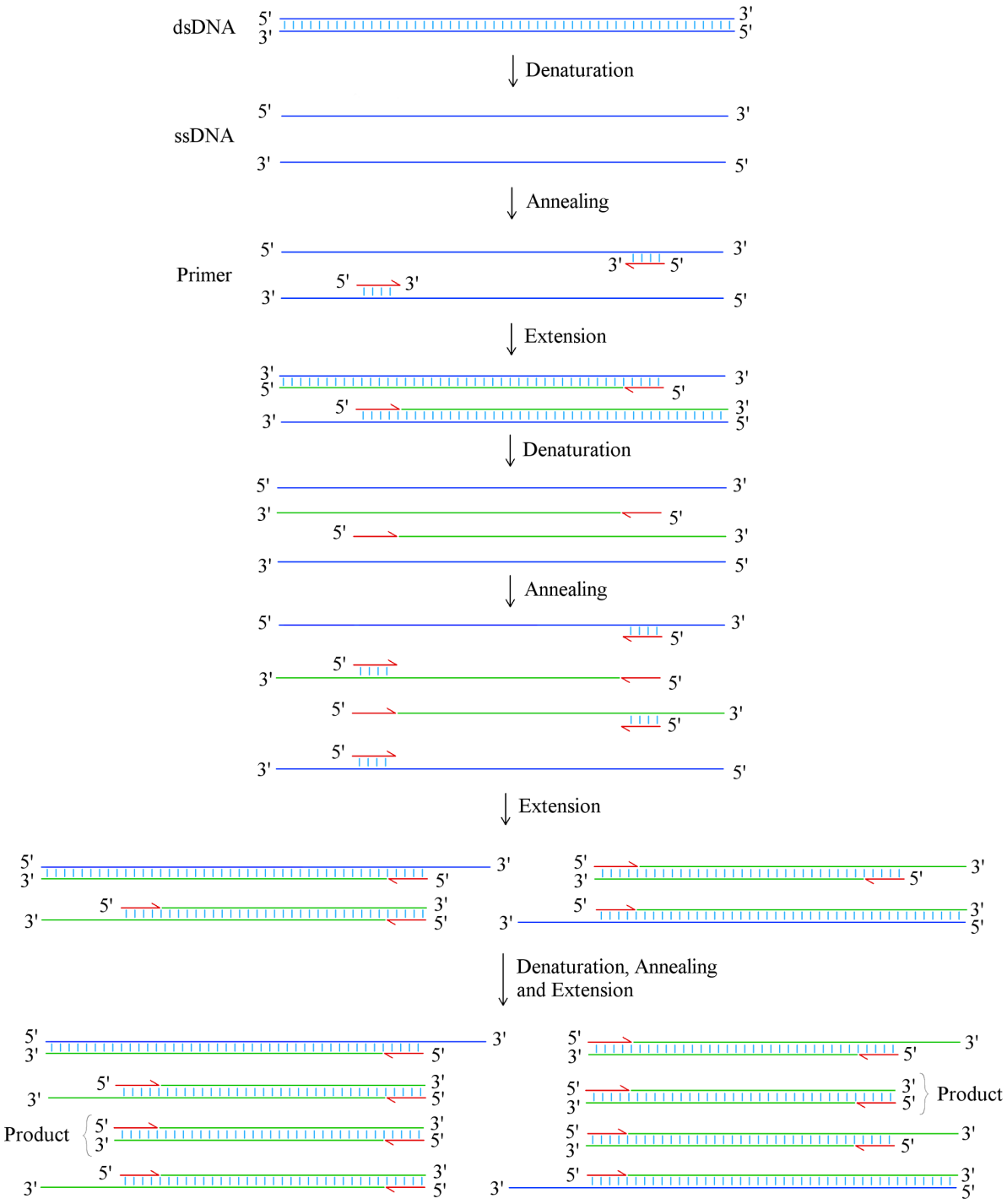


Figure 3: The basic concept of PCR involving multiple cycles of denaturation, annealing and extension. The figure shows amplification through the three first cycles. During denaturation the temperature is raised until the hydrogen bonds (light blue) between double stranded template DNA (dark blue) breaks, creating single stranded DNA. The temperature is then lowered, allowing primers (red) to bind to the complementary sequence of the template. The temperature is then slightly increased during the extension phase so that polymerase can synthesize new DNA-strands (green). [61]

The efficiency of a PCR-reaction is a measurement of how close the reaction is to a doubling of PCR product every cycle. The detection limit indicates how small the initial concentration of the DNA-template can be for a successful amplification to occur. If the DNA concentration of the initial sample is known, both efficiency and detection limit can be estimated by creating a serial dilution of the original sample and noting the Ct-values of each of the dilutions. The Ct-value is the number of cycles it takes for the fluorescence level to cross a user-defined threshold, see figure 4. As the sample becomes more diluted it will require more cycles to pass the threshold, eventually revealing what the lowest dilution that allows successful amplification is. Efficiency can be calculated by creating a standard curve of the Ct-values plotted against the initial amount of template DNA, and then using the slope in equation 1. Ideally, the standard curve should be linear with $R^2 > 0.980$, the efficiency should lie between 90 and 105 %, and there should be consistency across replicates in the dilution series [61, 63].

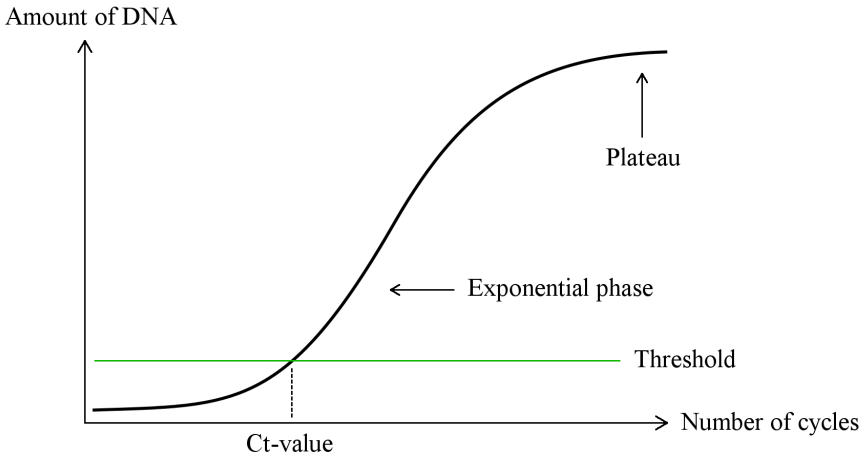


Figure 4: The amount of DNA created during PCR. There is an exponential increase in PCR product until it reaches a plateau where further amplification of product will be minimal. The number of cycles it takes for the fluorescense level to cross a threshold level is called the Ct-value.

$$\text{Efficiency} = 10^{-\frac{1}{slope}} - 1 \tag{1}$$

After real-time PCR performed using a dye such as EvaGreen, it is possible to do a melt analysis of the products to determine their exact melting temperature [64]. To do this, the temperature is gradually increased while doing multiple readings of the fluorescence signal. This will cause double-stranded DNA to split into single stranded DNA, whereby the bound dye will dissociate resulting in a sharp drop in fluorescence at the primer product melting temperature. It can easily be visualised by making a derivative plot, illustrated in figure 5. [61]

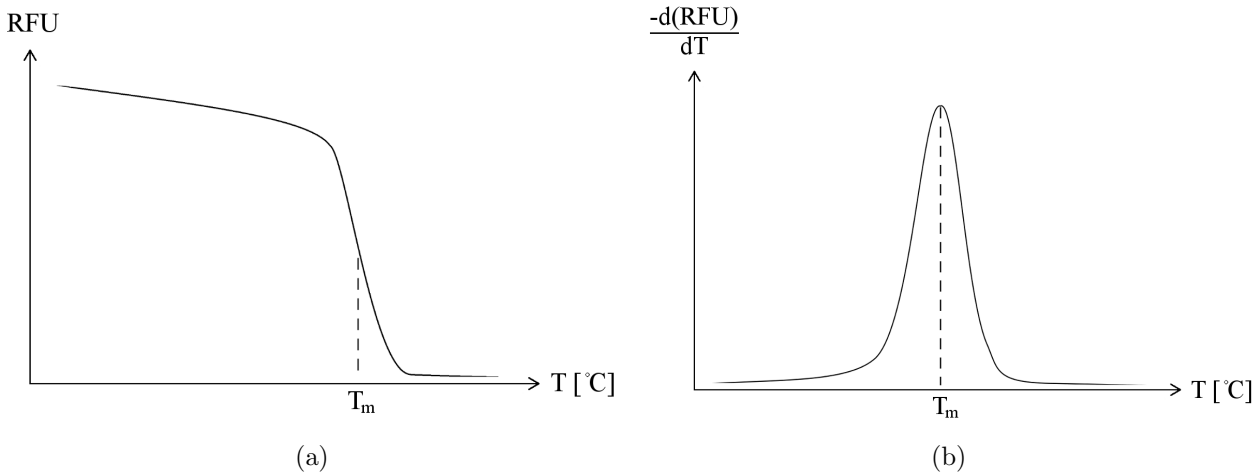


Figure 5: Melt analysis performed after real-time PCR. (a) The relative fluorescence of the PCR product (RFU) plotted against temperature reveals the melting temperature as a sharp drop in fluorescence. (b) is the derivative of plot (a).

1.4 High Resolution Melt Analysis

High resolution melt analysis (HRMA) is a melt profile with higher resolution than normal melt analysis [65]. This is made possible by increasing the temperature in smaller steps during the melt analysis, typically 0.2 °C, which enables differentiation of PCR-products based on subtle differences in their melting temperature, such as single nucleotide polymorphisms (SNPs) or small insertions and deletions [66]. There are four classes of SNPs: class I includes C>T, T>C, G>A, A>G SNPs, class II includes C>A, A>C, G>T, T>G SNPs, class III includes C>G, G>C SNPs, and class IV includes A>T, T>A SNPs [67]. Class III and IV type SNPs are harder to detect than class I and II because they do not alter the number of hydrogen bonds between the base pairs [65].

When doing HRMA it is recommended to use a saturating fluorescent dye in order to obtain a more reliable fluorescence signal. Unlike a non-saturating dye (such as SYBR), a saturating dye (such as EvaGreen) will bind to all available positions of the double-stranded DNA, making it better suited to distinguish single base pair differences [68]. The effect of chosen hardware and software for HRMA has been shown to vary greatly, and it seems that instruments made specifically for HRMA achieves more precise results [68, 69].

HRMA-compatible software must be used in order to properly process melt data. The software will carry out a normalisation of the fluorescence (y-axis) in pre- and post-melt regions to relative values 1 and 0 respectively in order to eliminate background fluorescence [70], see figure 6. The normalised data can then be used to create a differential plot. This plot is similar to the derivative plot but the curves of different PCR products are placed relatively to a reference curve, or a reference cluster of similar PCR products. By analysing PCR products with different types of SNPs in this way, they will form seperate clusters with distinct melting profiles which makes it possible to differentiate them.

It is recommended to analyse small PCR products when doing HRMA, as small nucleotide changes will have a larger effect on for instance, a 50 bp product than a 100 bp product [70].

A normalisation of both fluorescence (y-axis) and temperature (x-axis) produces a tempera-

ture shifted plot which may help to discriminate between products that produce differently shaped curves. Where a normal HRMA might not be able to differentiate between two of the PCR-products (blue and dark green curve in figure 6), a temperature shifted curve may be able to, see differential plot of figure 7. The drawback to this method is that any discriminatory power provided by differences in product melting temperature will be lost [71].

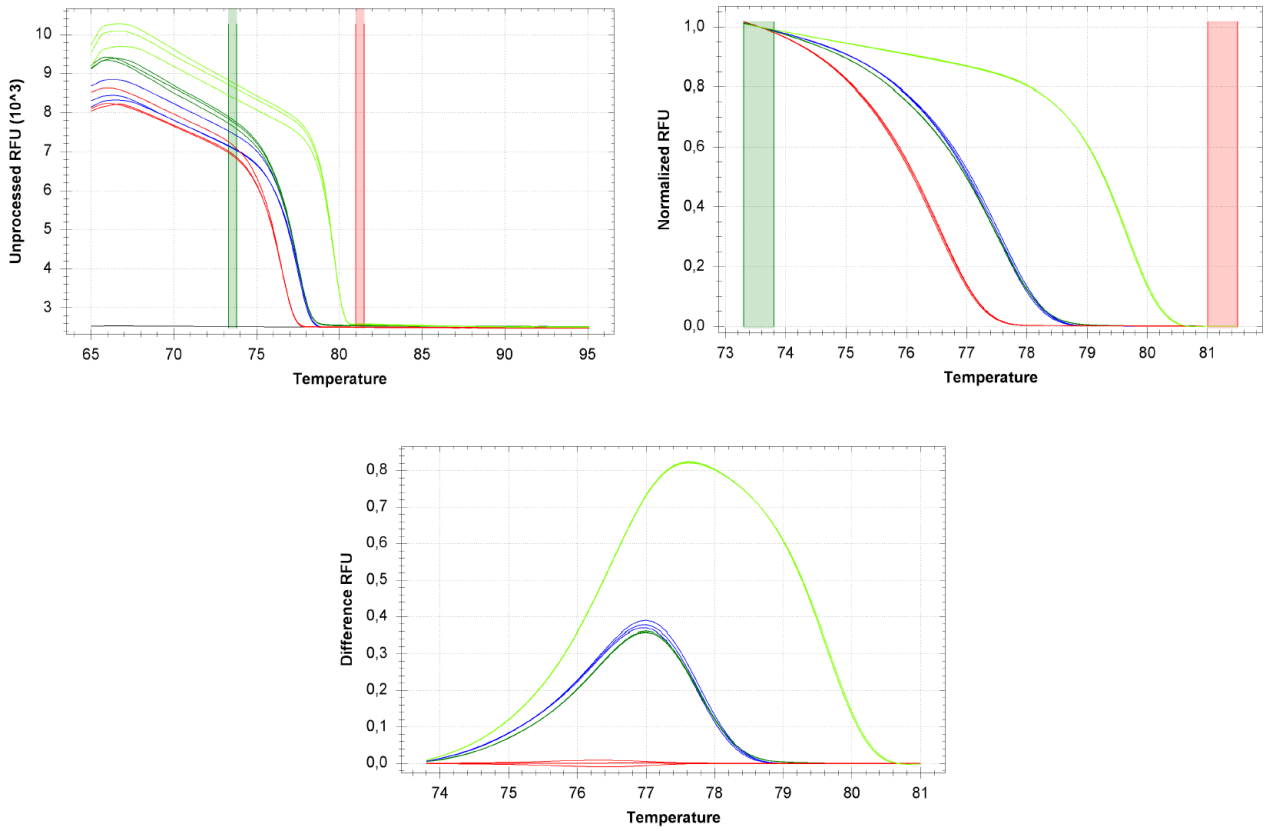


Figure 6: HRMA of real-time PCR products. Top left: A high resolution melt profile of real-time PCR-products with pre- and post-melt region shown as green and red vertical columns. Top right: A normalisation of the fluorescence data in the pre- and post-melt regions to relative values of 1 and 0. Bottom: A differential plot of the normalised melt data.

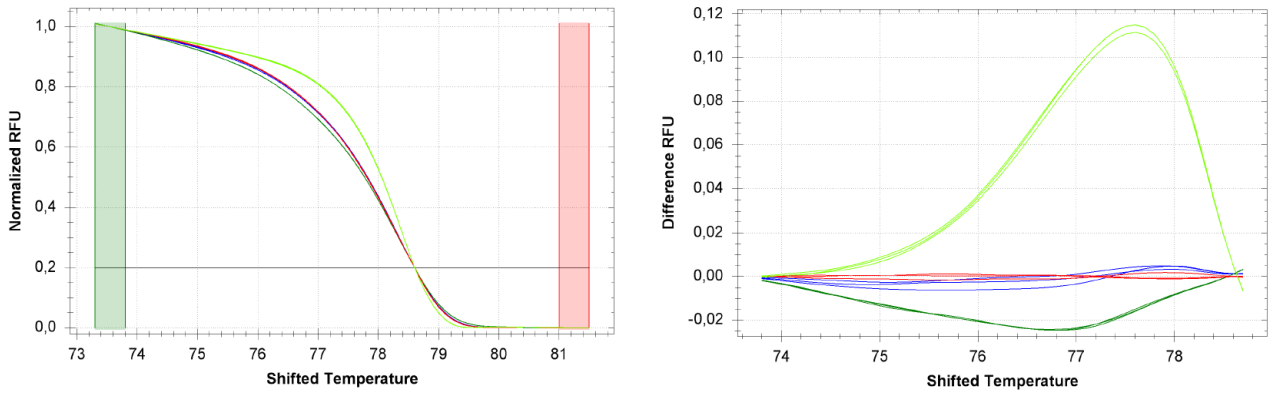


Figure 7: HRMA of real-time PCR products. Left: A normalisation has been carried out with respect to both fluorescence and temperature. Right: A differential plot of the normalised melt data.

1.5 Primer Design Considerations

A good primer is a primer that binds to its intended target with high specificity. The annealing temperature between the primer and the template strand, T_a, is generally estimated to be 5 °C lower than the melting temperature, T_m [61]. The T_m of the primer/template strand is dependent on the length and GC-content of the primer, and can be calculated in a multitude of different ways. A long and GC-rich primer will have a higher melting temperature, while a short and AT-rich primer will have a lower one. A high annealing temperature will help the primer be more specific because it lowers the number of similar places on the template strand it can stably bind to at that temperature. The T_a of each of the primers in a pair should be within ± 5 °C of each other to ensure that both will be able to bind to the target at the chosen annealing temperature [61].

To ensure that the primers are free to bind to the template, they should be designed without the ability to hybridise with themselves, forming hairpins, or each other, forming primer dimers [61]. Primers containing nucleotide repeats such as "ATATATGATAT" or runs such as "GGGGCGG" should be avoided because it can cause mispriming. [63]

Mismatches between the primer and the target region should be avoided because it will result in a lower amount of primers binding. It may also force one to lower the annealing temperature, increasing the risk of unspecific primer binding. A mismatch between the terminal nucleotide at the 3'-end of the primer and the template will not favour binding of polymerase [61]. The further away from the 3'-end of the primer the mismatch is, the less impact it will have on whether or not polymerase will be able to bind [72]. Targeting a conserved region is advisable to avoid mismatches [61].

Longer primers are more specific because they are less likely to bind to an unintended target. Specificity can be tested *in silico* by using NCBI's BLAST (basic local alignment search tool), which will indicate potential unwanted target regions for the primers.

2 Aim of the study

As the outbreak in 2006 in Germany showed, STEC can easily spread through a large population, and early detection is crucial to prevent outbreaks. As both conventional PCR and sequencing are cumbersome and relatively expensive, this project focuses on finding a more efficient method of subtyping Stx2 strains. Real-time PCR with HRMA was evaluated and chosen for this purpose because it is quick and cost-effective, does not require much personnel, and the equipment required for it is readily available in most labs. In detail, the project involves:

- A bioinformatic part where Stx2 nucleotide sequences are collected and edited.
- Analysis and identification of appropriate primer target regions and primer design.
- Evaluation of primers in conventional PCR and real-time PCR with HRMA using reference strains and a collection of Stx2 positive samples.

3 Materials and Methods

Two different sets of primer pairs were developed for the subtyping of Stx2, described in sections 3.3 and 3.4, each set consisting of three primer pairs. The three primer pairs in a set is run in three seperate real-time PCR tubes which contain the same Stx2 sample. The three resulting PCR products are then analysed with HRMA which will produce three different curves, or HRM profiles, that may be used to identify the Stx2 subtype.

3.1 Stx2 Nucleotide Sequences

DNA sequences for different Stx2 variants were found by browsing the nucleotide database at NCBI[1] and scientific literature. An article by Scheutz *et al* [29] listed a large number of different Stx2 variants they had used in their work to revise the Stx nomenclature. The sequences of these variants were considered relatively reliable in comparison to those found by searching for "Stx2" (and other related searchwords) at NCBI, and the accession numbers from [29] were used to gather sequence data from NCBI.

Editing and analyses of the DNA sequences were done using the JalView software[2]. The sequences were modified by removing everything that did not code for the Stx2 protein, except an intergenic sequence at nucleotide position 1137-1147 using accession z37725.1 (963–975 using the numbering provided after multiple alignment in this study). The exact position of the intergenic sequence may vary somewhat depending on the subtype and variant in question. Multiple alignment was performed using a feature in JalView that utilises Clustal Omega [73], and identical sequences were removed by using the option "remove redundant sequences". See appendix A for a full list of accession numbers used. Any references to positions hereafter will be done use the numbering system provided by multiple alignment unless otherwise specified.

3.2 Extracted DNA from Bacterial Strains and Mixed Cultures

DNA was extracted from *E. coli* reference strains listed in table 1 provided by Flemming Scheutz and Statens Serum Institut. The DNA extraction was kindly performed by the Department of microbiology at St. Olavs Hospital using NucliSENS easyMAG. DNA concentrations were measured using NanoDrop 1000 Spectrophotometer.

Table 1: Reference strains for each Stx2 subtype used in the development and testing of primers.

Variant	Strain	NCBI accession number
Stx2a-O48-94C	94C	Z37725.1
Stx2b-O118-EH250	EH250	AF043627.1
Stx2c-O174-031	031	L11079.1
Stx2d-O73-C165-02	C165-02	DQ059012.1
Stx2e-O139-S1191	S1191	M21534.1
Stx2f-O128-T4-97	T4/97	AJ010730.1
Stx2g-O2-7v	7v	AY286000.1

[1] http://www.ncbi.nlm.nih.gov/nuccore/
[2] http://www.jalview.org/

The number of *E. coli* genomes per µL of sample were calculated using equations 2–4 and used in estimation of detection limit. First, the molar mass of the M_{genome} [g/mol] of the *E. coli* genome was calculated:

$$M_{genome} = M_{nucleotide} \cdot Number\ of\ nucleotides\ per\ genome \tag{2}$$

Where $M_{nucleotide}$ is the molecular weight per nucleotide which was set to 660 [g/mol] and *Number of nucleotides per genome* was set to 5528445 using GenBank accession AE005174.2 as a reference. Then the number of moles, n, in 1 µL was calculated:

$$n = \frac{m}{M_{genome}} \tag{3}$$

Where m is the the mass [g] of *E. coli* genomes in 1 µL. Finally, the *Number of genomes per* μL was calculated:

$$Number\ of\ genomes/\mu L = n \cdot N_A \tag{4}$$

Where N_A is Avogadro's number $6.0221413 \cdot 10^{23}$.

Extracted DNA from STEC isolates and mixed cultures from patient stool samples were kindly provided for the study by the Department of Medical Microbiology, St. Olavs Hospital, see tables 2 and 3. No information from the patients was available, only that that the culture contained Stx2, or pure culture of STEC with Stx2, and in most cases which subtype.

Table 2: List of STEC isolates containing Stx2 available from St. Olavs from the time period 1996-2010. Subtyping had been performed by the Department of Foodborne Infections, Norwegian Institute of Public Health, Oslo.

Subtype	Sample name
Stx2a	s26, s30, s34, s36, s39, s48
Stx2c	s25, s29, s33, s37, s40, s44
Stx2d	s21, s27, s41, s42, s47
Stx2b	s22, s31, s35, s38, s43, s46
Stx2e	s45
Stx2g	s22a
Stx2a + Stx2c	s23, s24, s28, s32

Table 3: List of extracted DNA from mixed culture fecal samples gathered from patients at St. Olavs Hospital where Stx2 had been detected by real-time PCR during the time period 2012-2014. Subtype classifications had been done at the Department of Foodborne Infections, Norwegian Institute of Public Health, Oslo.

Subtype	Sample name
Stx2a	s1, s6, s12, s16, s49, s50
Stx2c	s2, s13, s15, s17, s51, s52
Stx2b	s9, s10, s14, s18
Stx2g	s19
Stx2a + Stx2c	s3, s4, s5
Stx2a + Stx2d	s7

3.3 Primer Set 1

Three primer pairs were designed using Oligo7: acdF1-R1, begF2-R2, and fF1-R1. Special consideration was made with respect to T_m to ensure they were approximately the same for all the primers.

The target subtypes of primer pair acdF1-R1 were Stx2a, Stx2c, and Stx2d. The targets of begF2-R2 were Stx2b, Stx2e, and Stx2g. The target of fF1-R1 was Stx2f. Degenerate primers were ordered to make sure they could bind to all the strains belonging to their targeted subtypes.

3.3.1 Conventional PCR

Conventional PCR was carried out with a Bio-Rad T100 Thermal Cycler in a total reaction volume of 50 µL containing 0.2 µL AmpliTaq Gold DNA Polymerase [5 U/µL] from life technologies, 5 µL Buffer I with $MgCl_2$ [10 x] which was supplied with the polymerase, 2.5 µL dNTPs [1 mM], 35.3 µL molecular grade water, 2.5 µL of each of forward and reverse primer [10 µM], and 2 µL DNA template diluted 1:10 with water. The PCR thermocycler conditions are shown in table 4.

Conventional PCR and agarose gel electrophoresis were performed in order to verify that the number of PCR products and sizes were correct. All of the primer pairs were tested on each of the reference strains listed in table 1. Optimum T_a was estimated for primer pair acdF1-R1 on reference strain Stx2a and for primer pair begF2-R2 on Stx2b by making a gradient from 48–65 °C.

Table 4: Conventional PCR thermocycler conditions used with primer pairs acdF1-R1, begF2-R2, and fF1-R1.

		Temperature [°C]	Time [minutes]
	Activation	94	10
35 cycles	Denaturation	94	1
	Annealing	51.3	1
	Extension	72	2
	Extra elongation	72	7

3.3.2 Real-time PCR and HRMA

Real-time PCR with HRMA was carried out using BioRad CFX96 in a total reaction volume of 20 µL containing 10 µL SsoFast EvaGreen Supermix from Bio-Rad, 2 µL of each of forward and reverse primer [5 µM], and 2 µL DNA template diluted 1:10 with water. The thermocycler conditions are shown in table 5. Real-time PCR with HRMA was performed in order to test whether it was possible to distinguish the reference strains from each other. All of the primer pairs were tested on each of the reference strains and evaluated using HRMA. The detection limit and efficiency were calculated from a triplicate 10-fold dilution series for reference strains Stx2a and Stx2b using primer pairs acdF1-R1 and begF2-R2 respectively. The detection limits and efficiencies were calculated in the same manner as the following examples show. The detection limit at dilution $1{:}10^4$ using 2 µL template DNA is calculated using the answer from equation 4 in equation 5:

$$\text{Detection limit} = \frac{\textit{Number of genomes per } \mu L \cdot 2}{10^4} \tag{5}$$

Efficiency is calculated by first making a standard curve of the Ct-values plotted against the dilutions of template DNA:

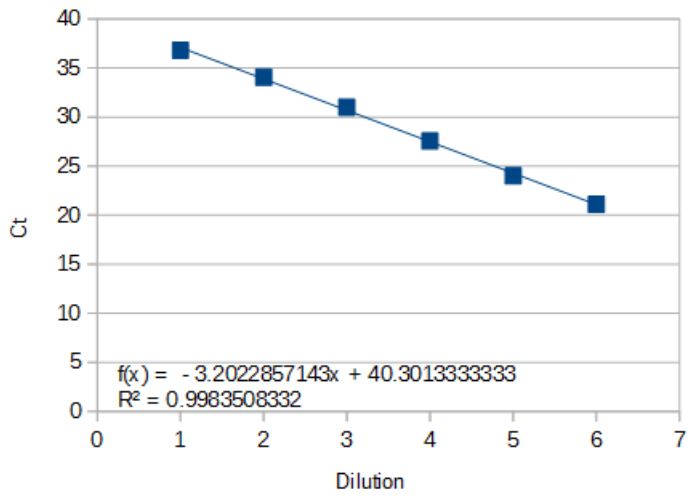


Figure 8: A standard curve of Ct-values plotted against the dilutions of template DNA shown with the equation of the curve and the coefficient of determination, R^2.

Using the slope of the standard curve (figure 8) with equation 6 reveals the efficiency as:

$$\text{Efficiency} = 10^{-\frac{1}{-3,20}} - 1 \cdot 100\% = 105\% \tag{6}$$

The primers were finally tested at T_a=54.5 °C on pure culture samples containing Stx2 (see table 2), and their HRM profiles evaluated to see whether samples of the same subtype formed similar HRM profiles.

Table 5: Real-time PCR and HRMA thermocycler conditions used when testing primer pairs acdF1-R1, begF2-R2, and fF1-R1.

		Temperature [°C]	Time
39 cycles	Activation	98	2 min
	Denaturation	98	0.05 sec
	Annealing/Extension	54.5	0.05 sec
	Melt curve	65-95 °C in 0.02 increments	0.05 sec/increment

3.4 Primer Set 2

The first set of primers (section 3.3) were not able to discriminate between all Stx2 subtypes, and therefore alternate primers were made: F1-R1, F3-R2, and F4-R4. These primers were designed to amplify smaller products, making differentiation of subtypes with HRMA easier.

The target subtypes of primer pair F1-R1 were Stx2a, Stx2c, Stx2d, Stx2e, and Stx2g. The targets of F3-R2 were Stx2c, Stx2d, and Stx2b. The target of F4-R4 was Stx2f.

3.4.1 Conventional PCR

As with the first primer pair set, conventional PCR and agarose gel electrophoresis were performed with the new primer pairs as well. All of the primer pairs were tested on each of the Stx2-subtypes. Reagents and protocol were the same as those used with the first primer pair set in conventional PCR, except for T_a which was 52 °C.

3.4.2 Real-time PCR and HRMA

Reagents and protocol were the same as those used with the first primer pair set in real-time PCR, except for T_a. Primer pairs F1-R1 and F3-R2 were tested on each of the reference strains and evaluated using HRMA at T_a = 52 and 56 °C. Primer pair F4-R4 was tested on all the reference strains only at T_a = 56 °C. A temperature gradient ranging from 56-66 °C was made to determine the optimum T_a for the primers on the reference strains that were amplified at 56 °C.

The detection limit and efficiency were calculated at 60 °C (optimum annealing temperature) on a triplicate 10-fold dilution series for reference strains Stx2a and Stx2g using primer pair F1-R1, on reference strains Stx2c and Stx2b using primer pair F3-R2, and on reference strain Stx2a using primer pair F4-R4. The PCR protocol were modified to run 59 cycles instead of 39. Calculations were done in the same manner as with the first primer set.

The primers were tested at T_a=60 °C on pure culture samples containing Stx2 (see table 2), and their HRM profiles evaluated to see whether samples of the same subtype formed similar HRM profiles. To see how various DNA-template concentrations affected amplification and HRM profiles, the primers were also tested on mixed culture samples (see table 3).

4 Results

4.1 Sequence Data

Analysis after multiple alignment revealed that *stx2a*, *stx2c*, and *stx2d* showed high sequence identity with each other. *stx2b*, *stx2e*, and *stx2g* showed lower sequence identity with each other and with the other subtypes, and *Stx2f* was very dissimilar to all the other subtypes.

4.2 Template Concentrations

Table 6 lists DNA concentrations of the reference strains listed in table 1, as well as the calculated number of genomes per µL.

Table 6: Average measure of DNA-template concentrations in Stx2 reference strains and approximate number of genomes in each sample.

Subtype	Stx2a	Stx2b	Stx2c	Stx2d	Stx2e	Stx2f	Stx2g
Average concentration [ng/µL]	21.0	20.5	21.2	15.5	11.7	19.8	23.4
10^6 genomes/µL	34.7	33.8	35.0	25.6	19.3	32.7	38.6

4.3 Primer Set 1

Designed primers with T_m provided by Oligo7 and their target regions are presented in table 7. acdF1-R1 target subtypes Stx2a, Stx2c, and Stx2d. begF2-R2 targets Stx2b, Stx2e, and Stx2g. fF1-R1 target Stx2f. The three primer pairs target regions which result in three differently sized products with different T_ms. This should make it easy to separate subtypes Stx2a, Stx2c, Stx2d from Stx2b, Stx2e, Stx2g from Stx2f in HRMA.

Table 7: Sequences, T_m, position and PCR product lengths of primers acdF1-R1, begF2-R2 and fF1-R1 which were used to amplify Stx2 subtypes.

Primer	Sequence 5' → 3'	T_m [°C]	Position	PCR product length [bp]
acdF1	TAATACAGCWGCAGCGTTTC	56.4	903-922	245
acdR1	GTRACGGTYGCAGATTCCAG	56.5	1147-1128	
begF2	ATGGGAAAGTAATACMGCAG	56.0	894-913	126
begR2	ACAGAAACMAAWGCAAATAAARCC	57.1	1019-996	
fF1	AATACCTTTACTGTGAGGGTG	56.0	1081-1101	62
fR1	GGCTGCAAATTCCATCTGTTC	57.3	1142-1122	

Figure 9 shows each of the reference strains used to design the primers along with the primer target regions. Note that this is an excerpt of sequences from multiple alignment performed on a large number of strains listed in appendix A, and the consensus sequence in figure 9 is a result of multiple alignment using all these strains.

The nucleotide sequences amplified with primer pair acdF1-R1 were highly conserved in Stx2a, Stx2c, and Stx2d, yet contained SNPs uniqely conserved only in some of the subtypes. The nucleotides marked orange in figure 9 are higly conserved SNPs in subtypes Stx2a and Stx2c. The

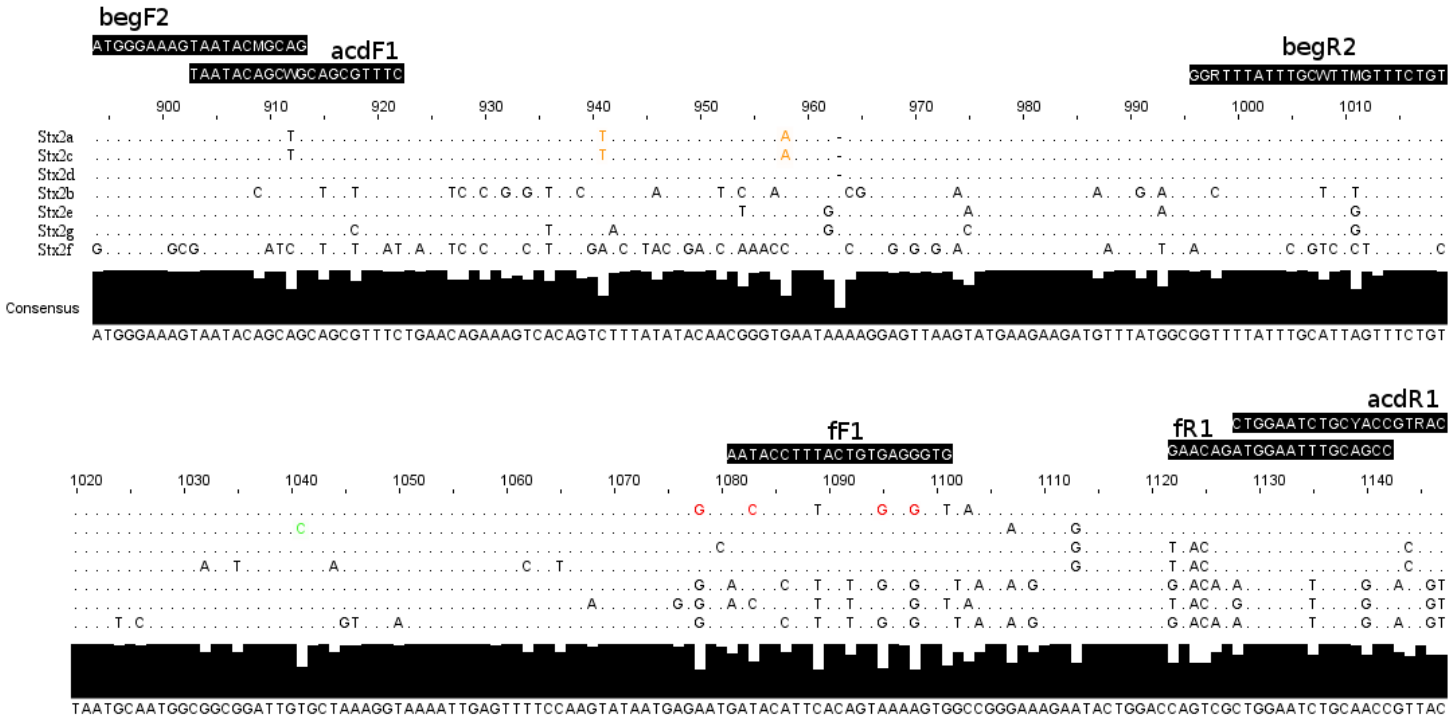


Figure 9: Representatives of each Stx2-subtype along with primer binding sites for primer pairs acdF1-R1, begF2-R2 and fF1-R1. The presented sequences are from the reference strains listed in table 1. Note that this is an excerpt of sequences from multiple alignment performed on a large number of strains listed in appendix A, and the consensus sequence is a result of multiple alignment using all these strains. The coloured nucleotides are SNPs which are highly conserved in the corresponding strains.

nucleotide marked green is conserved in Stx2c, and those marked red are conserved in Stx2a. These regions were chosen on the basis that HRMA would be able to differentiate between the subtypes, even though there were multiple SNPs involved and the PCR products would be rather large. The same type of reasoning was employed when trying to find suitable target regions for the two other primer pairs.

4.3.1 Conventional PCR

Primer set 1 was tested in conventional PCR on the reference strains listed in table 1 to determine if the number and sizes of the PCR products were correct. The primer pairs were tested as indicated in table 8 and the gel results are presented in figure 10. acdF1-R1 produced the desired 245 bp product from Stx2a, Stx2c and Stx2d. begF2-R2 was able to amplify the desired 126 bp product from Stx2b, Stx2e and Stx2g. Primer pair fF1-R1 was barely able to amplify the expected 62 bp product from Stx2f (marked with red circle), and further testing with this primer was discontinued.

acdF1-R1 and begF2-R2 were then tested on *all* of the Stx2 reference strains as indicated in table 9, and the gel results are presented in figure 11. Primer pair acdF1-R1 produced 245 bp products in addition to a product of about 1000 bp for Stx2g. Primer pair begF2-R2 produced 126 bp products.

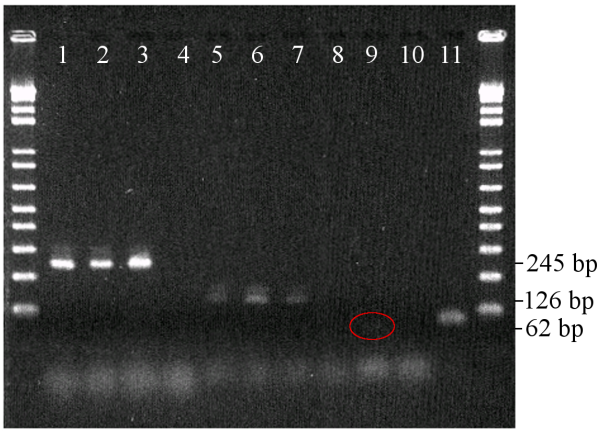


Figure 10: Agarose gel electrophoresis performed after conventional PCR as indicated in the adjacent table.

Table 8: Setup for testing of primer pairs acdF1-R1, begF2-R2, and fF1-R1 in conventional PCR on indicated Stx2 subtype reference strains.

Primer pair	Well	Sample	Amplification?
acdF1-R1	1	Stx2a	+
	2	Stx2c	+
	3	Stx2d	+
	4	N	-
begF2-R2	5	Stx2b	+
	6	Stx2e	+
	7	Stx2g	+
	8	N	-
fF1-R1	9	Stx2f	+ (very weak)
	10	N	-
acdF1-R1	11	Pos	+

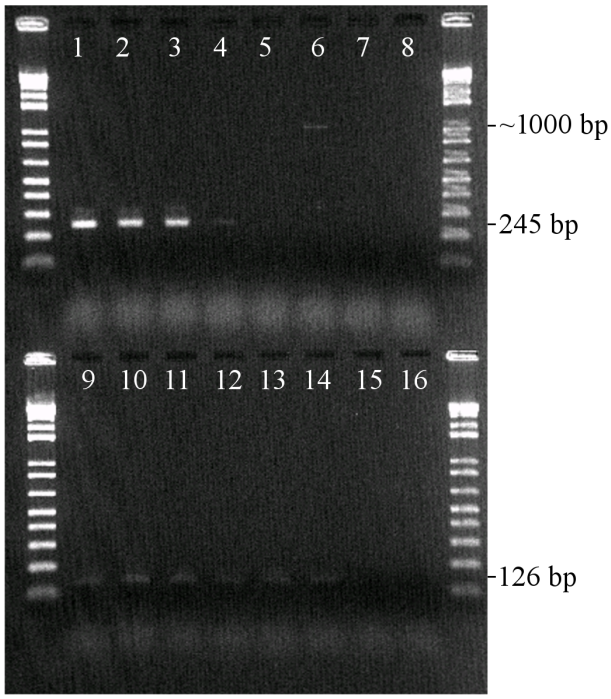


Figure 11: Agarose gel electrophoresis performed after conventional PCR as indicated in the adjacent table.

Table 9: Setup for testing of primer pairs acdF1-R1 and begF2-R2 in conventional PCR on indicated Stx2 subtype reference strains.

Primer pair	Well	Sample	Amplification?
acdF1-R1	1	Stx2a	+
	2	Stx2c	+
	3	Stx2d	+
	4	Stx2b	+ (very weak)
	5	Stx2e	-
	6	Stx2g	+ (very weak)
	7	Stx2f	-
	8	N	-
begF2-R2	9	Stx2a	+
	10	Stx2c	+
	11	Stx2d	+
	12	Stx2b	+
	13	Stx2e	+
	14	Stx2g	+
	15	Stx2f	-
	16	N	-

4.3.2 Real-time PCR and HRMA

The primer pairs were tested in real-time PCR on target reference strains as indicated in table 10, and the HRMA results are shown in figure 12. Stx2a, Stx2c, and Stx2d are clearly distinguishable

Table 10: Test setup for real-time PCR with primer pairs acdF1-R1 and begF2-R2 on indicated reference strains. Each subtype were tested in separate reactions.

Primer pair	Sample
acdF1-R1	Stx2a
	Stx2c
	Stx2d
begF2-R2	Stx2b
	Stx2e
	Stx2g

from Stx2b, Stx2e, and Stx2g. Discrimination between Stx2a and Stx2c was not possible, and neither was discrimination between Stx2b and Stx2g. A temperature shifted differential plot made it possible to distinguish Stx2a from Stx2c and Stx2b from Stx2g. By combining observations from the normal HRM differential plot with those from the temperature shifted one, it was possible to differentiate between the tested subtypes.

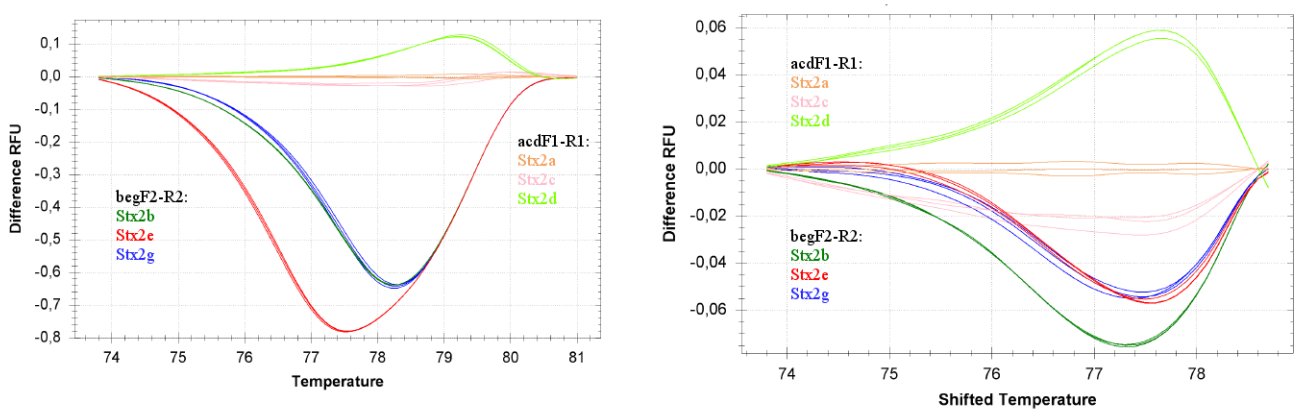


Figure 12: Left: A differential plot in HRMA made after real-time PCR of subtypes Stx2a, Stx2c, and Stx2d amplified with primer pair acdF1-R1 and subtypes Stx2b, Stx2e, and Stx2g with primer pair begF2-R2. Each reference strain was amplified with the primer pairs in seperate PCR tubes. Reference cluster: Stx2a with acdF1-R1. Right: A temperature shifted differential plot in HRMA made after real-time PCR of subtypes Stx2a, Stx2c, and Stx2d amplified with primer pair acdF1-R1 and subtypes Stx2b, Stx2e, and Stx2g with primer pair begF2-R2. Each reference strain was amplified with the primer pairs in seperate PCR tubes. Reference cluster: Stx2a with acdF1-R1.

The two primer pairs were then tested on *all* of the Stx2 reference strains in the same fashion as with conventional PCR (table 9), and the result is presented in figure 13. Primer pair acdF1-R1 amplified all reference strains except Stx2f, and begF2-R2 amplified all the strains. Combined with the temperature shifted plot described before, it would seemingly be possible to differentiate between all the subtypes.

The detection limit of primer pair acdF1-R1 on reference strain Stx2a was 6931 genome copies at a dilution of 10^4. The efficiency was calculated to be 108 % and 97 % with and without

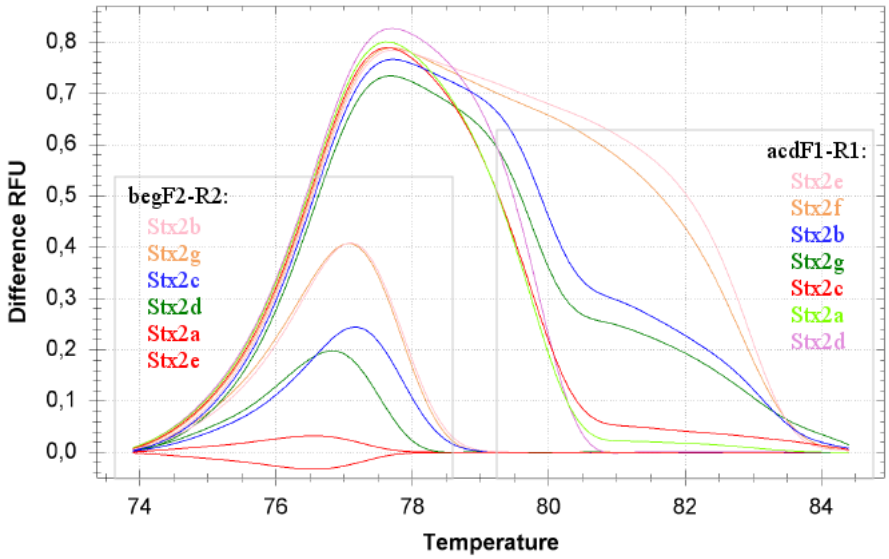


Figure 13: A differential plot in HRMA made after real-time PCR of all Stx2 reference strains amplified with primer pairs acdF1-R1 and begF2-R2. Each reference strain was amplified with the primer pairs in seperate PCR tubes. Reference cluster: Stx2e with begF2-R2.

the undiluted sample respectively. For primer pair begF2-R2 the detection limit was 5116 genome copies at a $1 : 10^4$ dilution with reference strain Stx2b, and the efficiency was 115% and 98 % with and without the undiluted sample respectively. The consistency across replicates was high, and the lowest coefficient of determination, R^2, was 0.98.

The primers were finally tested on a small number of pure culture samples (table 2) to see whether samples of the same subtype formed distinctive clusters which could be separated from clusters containing other subtypes. The results are presented in figure 14. It was possible to distinguish between the products made by the two different primer pairs, but samples belonging to the same subtype did not form distinctive clusters, and there was no real chance of discriminating between them. A temperature shifted curve did not improve the situation.

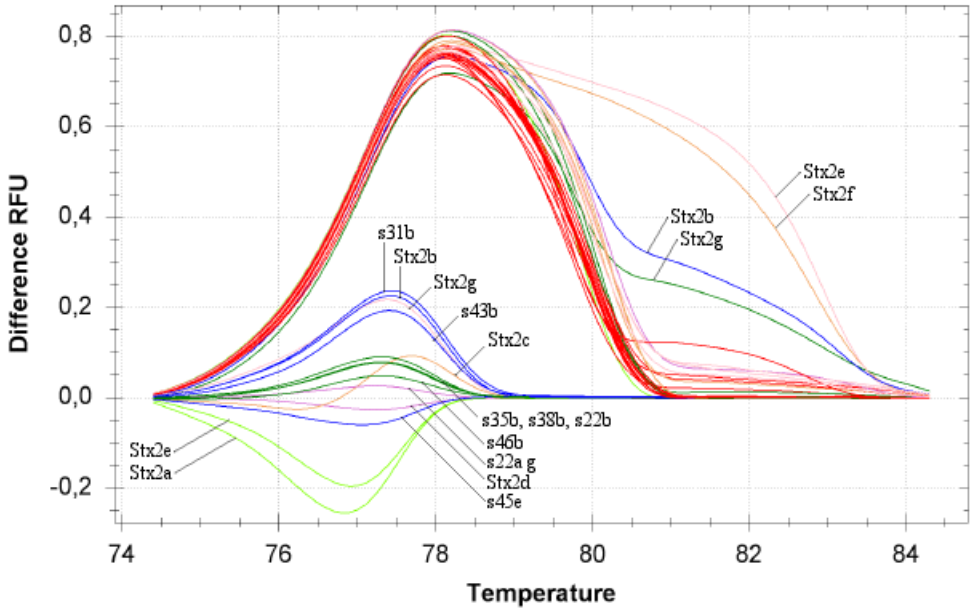


Figure 14: An HRMA differential plot made after real-time PCR of all Stx2 reference strains and the samples listed in table 2 with primer pairs acdF1-R1 and begF2-R2. The reference strains have been marked with Stx2 subtype and the samples have been marked with sample number and a letter indicating what subtype they belong to. Note that sample s22a is of subtype Stx2g, not Stx2a. The unmarked curves are the samples amplified with begF2-R2 (not marked because they are indistinguisable). Reference cluster: Stx2e with begF2-R2.

4.4 Primer Set 2

New primer were designed and their T_m (provided by Oligo7) and target regions are presented in table 11. Figures 15-17 show each of the reference strains used to design the primers along with primer target regions and coloured SNPs of critical importance for HRMA. The sequences are an excerpt from multiple alignment performed on a large number of strains listed in appendix A, and the consensus sequence is a result of multiple alignment with all these strains.

Table 11: Sequences, T_m, position and PCR product lengths of primers F1-R1, F3-R2 and F4-R4 which were used to amplify Stx2 subtypes.

Primer	Sequence 5' → 3'	T_m [°C]	Position	PCR product length [bp]
F1	GAACAGAAAGTCACAGTTTTTATATACAA	54.5	924-952	68 and 69
R1	ATAAACATCTTCTTCATACTTAACTCCT	56.8	992-965	
F3	TTAGTTTCTGTTAATGCAATGGC	53.7	1009-1031	97
R2	CGGCCACTTTTACTGTGAATGTA	58.6	1105-1083	
F4	CAACCAGAATGTCAGATAAC	52.9	838-857	265
R4	CAACCTTCACTGTAAATGTG	54	1102-1083	

Primer pair F1-R1 targets a region which contain a class I SNP at position 958, see figure 15 (SNP marked red), separating almost all Stx2a strains and all Stx2c strains from Stx2d. There

is a deletion at 963 (green arrow) for Stx2a, Stx2c and Stx2d, making it possible to distinguish Stx2a, Stx2c, and Stx2d from other subtypes (See appendix B for a full overview of conserved nucleotides in all strains of Stx2a, Stx2c, and Stx2d.). The region also contains a class II SNP at position 954 (orange), making it possible to separate Stx2e from Stx2g. Primer F1 was specifically designed so that the 3'-end would mismatch with Stx2b and Stx2f, making it unable to amplify these sequences. The result of all this is that an HRMA will be able to differentiate between Stx2d, Stx2e, and Stx2g, but Stx2a and Stx2c will have identical HRM profiles.

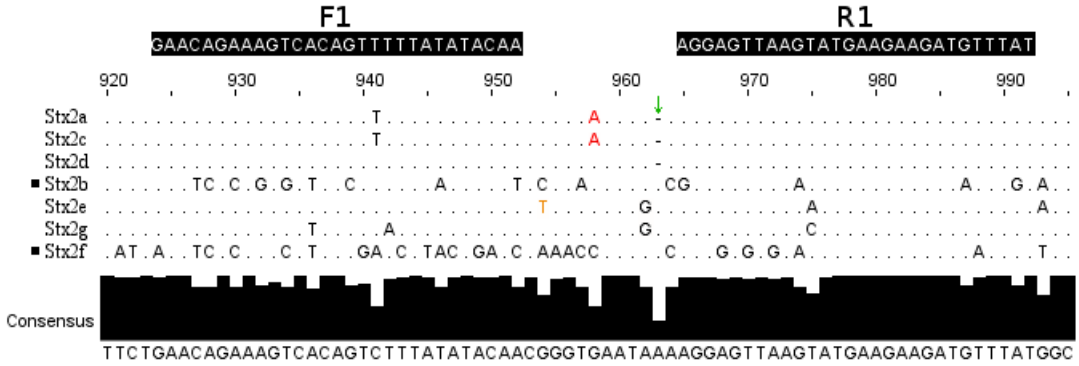


Figure 15: Representatives of each Stx2-subtype along with primer binding sites for primer pair F1-R1. The presented sequences are from the reference strains listed in table 1. Note that this is an excerpt of sequences from multiple alignment performed on a large number of strains listed in appendix A, and the consensus sequence is a result of multiple alignment using all these strains. The coloured nucleotides are SNPs which are highly conserved in the corresponding strains. The black squares mark the subtypes which should not be amplified.

Primer pair F3-R2 targets a region which contain a class I SNP at position 1041, see figure 16 (SNP marked orange), which is perfectly conserved in Stx2c but only partially conserved in Stx2a and Stx2d, see appendix B. The fact that the position is not fully conserved in Stx2d strains is not a problem because primer pair F1-R1 is already able to separate Stx2c from Stx2d. Primer R2 was specifically designed so that it would not amplify Stx2a due to a 3'-end mismatch at position 1083. A sample containing Stx2a could create an HRM-profile indicative of Stx2d using primer pair F1-R1, but had it truly been Stx2d then this primer pair would have amplified it. Stx2b contain four class II SNPs and one class I SNP, resulting in a lower T_m than both Stx2c or Stx2d. So, in combination with primer pair F1-R1, this primer pair should be able to differentiate between Stx2c and Sx2d, and under stringent conditions not amplify Stx2a, Stx2e, Stx2g or Stx2f due to multiple mismatches. The problem with primer pair F1-R1 where Stx2a and Stx2c would create identical HRM profiles is solved when using this primer pair in combination with F1-R1. Lets say a sample contains Stx2c; HRMA of the first primer pair alone would have created an HRM profile that could be either Stx2a or Stx2c. When running real-time PCR with primer F3-R2 as well, there would not be any amplification had it been Stx2a, thus it would have to be Stx2c.

There was a problem left: if both Stx2a and Stx2c or both Stx2a and Stx2d is present in the same sample, it would not be possible to distinguish them from Stx2c using only F1-R1 and F3-R2. To resolve this issue primer pair F4-R4 was designed to be Stx2a-specific. Primer R4 was designed to amplify only Stx2a and Stx2g; these are the only subtypes which match the 3'-end of the primer, marked orange in figure 17. Primer F4 was designed to amplify everything *but* Stx2g due to a

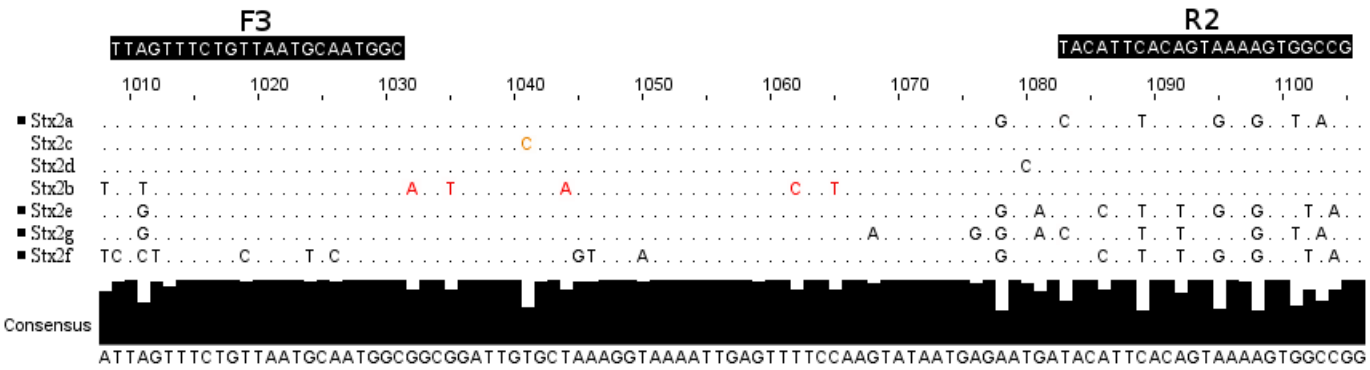


Figure 16: Representatives of each Stx2-subtype along with primer binding sites for primer pair F3-R2. The presented sequences are from the reference strains listed in table 1. Note that this is an excerpt of sequences from multiple alignment performed on a large number of strains listed in appendix A, and the consensus sequence is a result of multiple alignment using all these strains. The coloured nucleotides are SNPs which are highly conserved in the corresponding strains. The black squares mark the subtypes which should not be amplified.

mismatch in the 3'-end of the primer (marked red). Lets say that a sample contains both Stx2a and Stx2c. This would result in an HRM profile that is the same for both Stx2a and Stx2c when using primer pair F1-R1. With F3-R2 it would appear as being Stx2c. So it would be impossible to say whether this was a sample containing just Stx2c or both Stx2c and Stx2a. But if F4-R4 is used as well there would be an amplification which indicates that Stx2a is present as well, hence it cannot be only Stx2c.

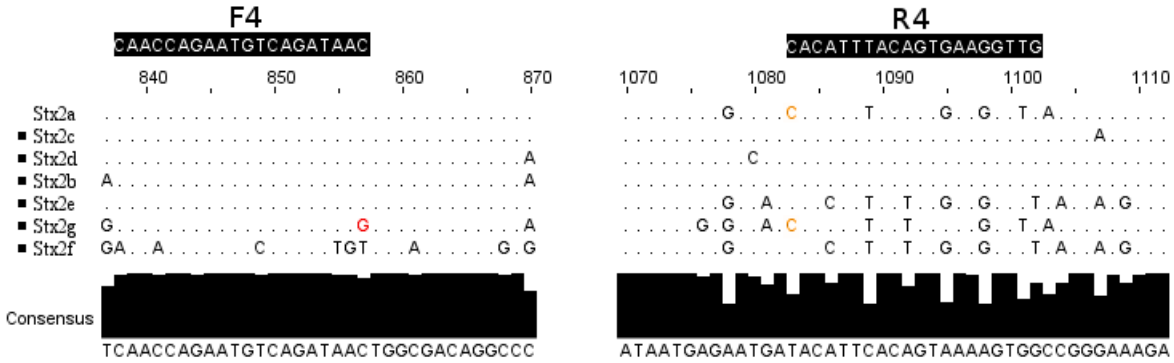


Figure 17: Representatives of each Stx2-subtype along with primer binding sites for primer pair F4-R4. The presented sequences are from the reference strains listed in table 1. Note that this is an excerpt of sequences from multiple alignment performed on a large number of strains listed in appendix A, and the consensus sequence is a result of multiple alignment using all these strains. The coloured nucleotides are SNPs which are highly conserved in the corresponding strains. The black squares mark the subtypes which should not be amplified.

By amplifying a sample with an unknown Stx2 subtype with the three primer pairs in three different real-time PCR tubes and then doing a HRMA, it should be possible to determine the subtype by evaluating the HRM profile of each of the three primer pairs.

4.4.1 Conventional PCR

Primer set 2 was tested on all reference strains listed in table 1 as indicated in table 12. Primer pair F1-R1 was able to amplify Stx2a, Stx2c, Stx2d, Stx2e and Stx2g, producing products of the expected sizes of 68 and 69 bp, see figure 18. Primer pair F3-R2 was able to amplify Stx2c, Stx2d and Stx2b, producing products of the expected size of 97 bp. Primer pair F4-R4 produced 265 bp products for Stx2a. Thus, all the primer pairs amplified their intended targets.

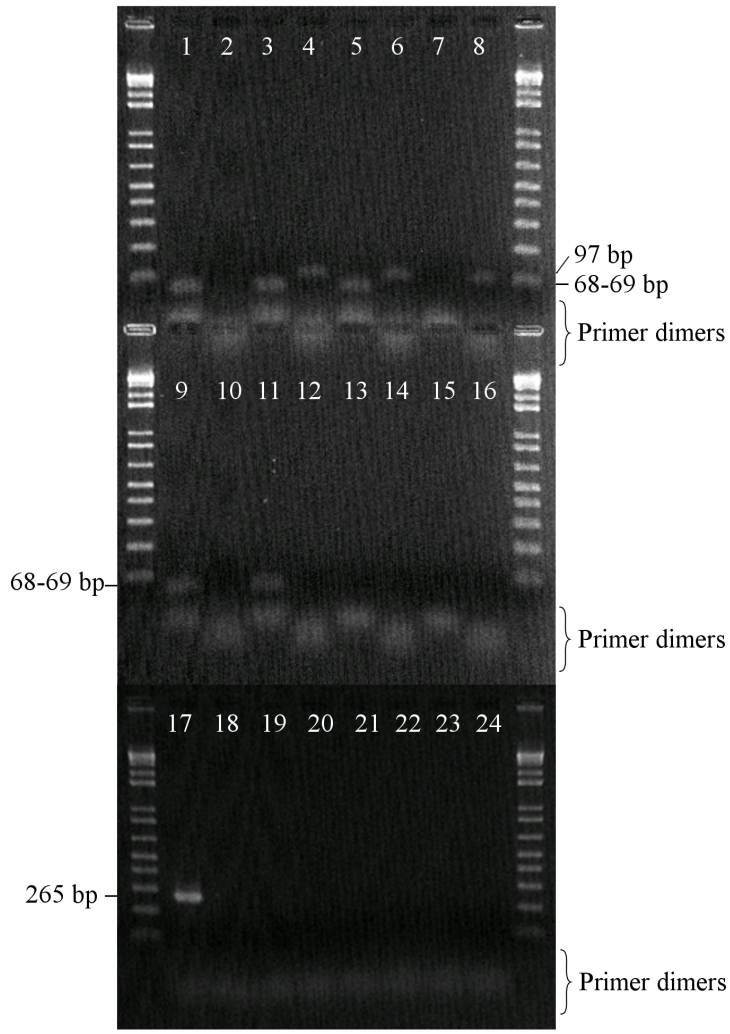


Figure 18: Agarose gel electrophoresis performed after conventional PCR as indicated in the adjacent table.

Table 12: Setup for testing of primer pairs F1-R1, F3-R2, and F4-R4 in conventional PCR on indicated Stx2 subtype reference strains.

Primer pair	Well	Sample	Amplification?
F1-R1	1	Stx2a	+
	3	Stx2c	+
	5	Stx2d	+
	7	Stx2b	-
	9	Stx2e	+
	11	Stx2g	+
	13	Stx2f	-
	15	N	-
F3-R2	2	Stx2a	-
	4	Stx2c	+
	6	Stx2d	+
	8	Stx2b	+
	10	Stx2e	-
	12	Stx2g	-
	14	Stx2f	-
	16	N	-
F4-R4	17	Stx2a	+
	18	Stx2c	-
	19	Stx2d	-
	20	Stx2b	-
	21	Stx2e	-
	22	Stx2g	-
	23	Stx2f	-
	24	N	-

4.4.2 Real-time PCR and HRMA

Primer pairs F1-R1 and F3-R2 were first tested on all of the Stx2 reference strains at $T_a = 52$ °C which produced some additional unwanted amplifications. The T_a was then raised to $T_a = 56$ °C to prevent these amplifications. F4-R4 was thereafter tested on all the reference stranis at $T_a =$

56 °C. At 56 °C there were no unwated amplifications for any of the primer pairs, and all of the targeted subtypes showed successful amplification. HRMA profiles at 56 °C is presented in figure 19.

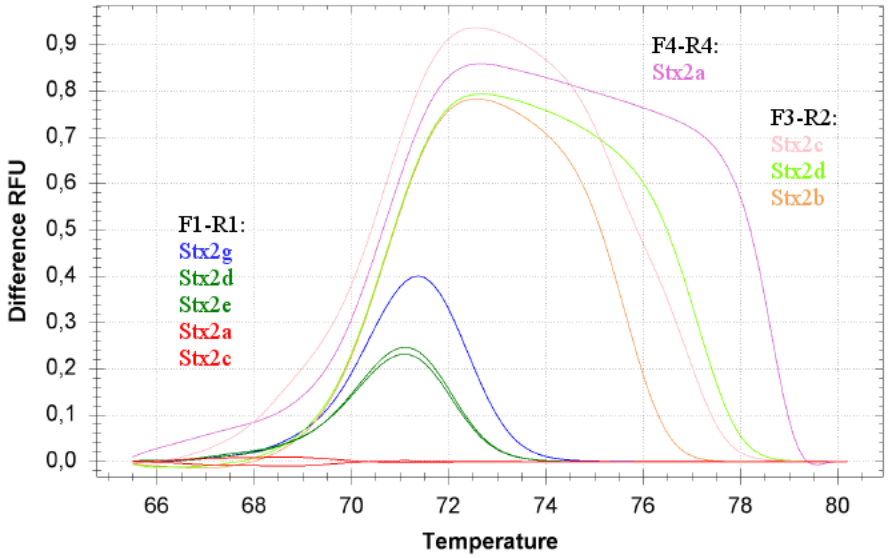


Figure 19: A differential plot in HRMA made after real-time PCR of all Stx2 reference strains amplified with primer set 2. Each reference strain was amplified with the primer pairs F1-R1, F3-R2, and F4-R4 in seperate PCR tubes. Reference cluster: Stx2a with F1-R1.

After these initial tests, the optimum T_a was determined to be 60 °C, and the HRM profiles were the same as those indicated in figure 19. Detection limits and efficiencies were then calculated after performing a 10-fold dilution series and are presented in table 13. The efficiencies of F4-R4 with Stx2a and F1-R1 with Stx2g were alarmingly high. Efficiencies in these two cases were much lower when the most concentrated sample was omitted, and even 127 % for F1-R1 with stx2g when omitting the two most concentrated samples. Efficiencies were generally lower for all the primer pairs when omitting the most concentrated sample. The consistency across replicates and coefficient of determination were poor for F1-R1 on Stx2a and F4-R4 on Stx2a. The consistency across replicates and coefficient of determination for the primer pairs were otherwise good, with a coefficient of determination above 0.99.

Table 13: Calculated detection limits and efficiencies after performing a 10-fold dilution series in real-time PCR on indicated reference strains and primer pairs.

Subtype & Primer pair	Detection limit	Efficiency [%]	Efficiency excluding undiluted sample [%]
Stx2a F1-R1	6932	105	105
Stx2g F1-R1	6536	580	148
Stx2b F3-R2	5116	106	97
Stx2c F3-R2	6767	105	99
Stx2a F4-R4	6932	161	127

Each of the primer pairs were tested on the reference strains and a small collection of pure

culture samples listed in table 2 in addition to sample s7 from the mixed culture (table 3) to see whether samples of the same subtype formed distinctive clusters which could be separated from clusters containing other subtypes. The HRMA results are presented in figure 20. With some exceptions (see below), all three primer pairs amplified their targets and produced HRMA profiles in line with those seen in figure 19, making it possible to subtype them.

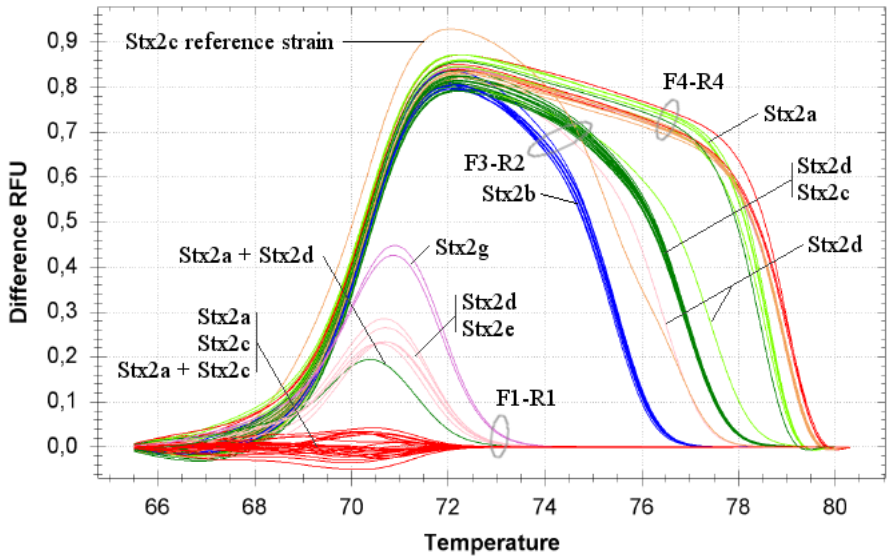


Figure 20: HRM profiles after amplification of the samples listed in table 2 in addition to reference strains listed in table 1 with primer pairs F1-R1, F3-R2, and F4-R4. Clusters have been marked with the subtypes they contain, excluding four exeptions (see text). Reference cluster: Stx2a with F1-R1.

The exceptions mentioned earlier are samples s41 and s21 (Stx2d), s45 (Stx2e), s44 (Stx2c), and s7 (Stx2a and Stx2d), although sample s7 produced an expected HRMA-profile upon a second try. Sample s41 clustered with Stx2a/Stx2c using primer pair F1-R1, Stx2d using F3-R2, and Stx2a using F4-R4. Sample s21 clustered with Stx2a/Stx2c using F1-R1, Stx2d using F3-R2, and showed very poor amplification with F4-R4 resulting in it not being picked up by the HRMA-software. s45 showed signs of late amplification using F1-R1, no amplification with F3-R2, and positive amplification using F4-R4 which is Stx2a specific. s44 was clustered with Stx2a/Stx2c using primer pair F1-R1, Stx2c using F3-R2, and Stx2a using F4-R3.

Each of the primer pairs were finally tested on the reference strains and a small collection of mixed culture samples listed in table 3 to see how different DNA template concentrations affected amplification and HRM profiles. Samples amplified with primer pairs F1-R1 and F3-R2 had Ct-values ranging from 17 to 30 with a threshold of 93, but the vast majority of samples had a Ct-value around 20. The samples amplified with primer pair F4-R4 had Ct-values evenly distributed from 20 to 40, showing poor amplification. Some samples where not amplified; sample s4 (Stx2a and Stx2c) and s19 (Stx2g) did not get amplified with any of the primers and sample s12 (Stx2a) showed very poor amplification which caused the HRMA software to exclude it. HRMA results are presented in figure 21. All of the samples showed HRM profiles expected for their respective subtype.

The Stx2c reference strain did not show the same HRMA profile as any of the Stx2c samples tested this far, see figures 20 and 21. It also formed two melt peaks in a normal melt analysis, see figure 22.

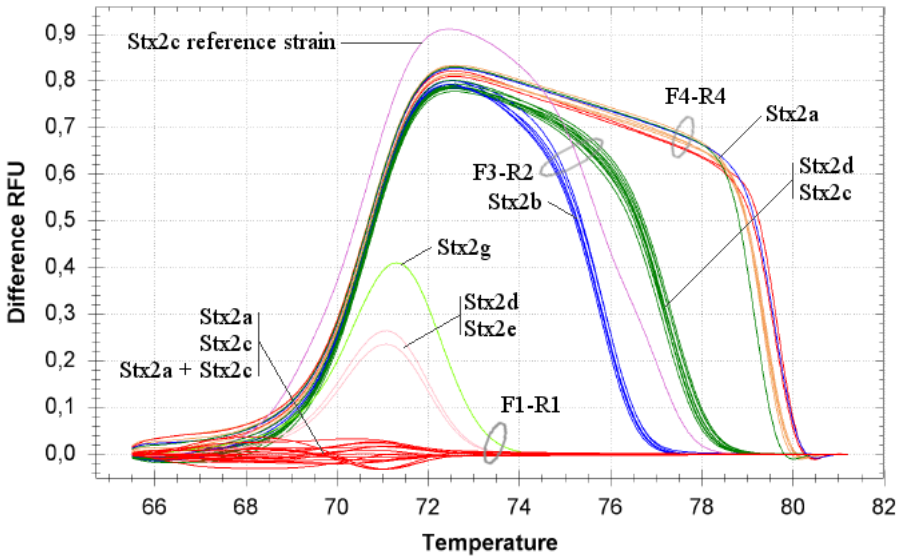


Figure 21: HRM profiles after amplification of the samples listed in table 3 in addition to reference strains listed in table 1 with primer pairs F1-R1, F3-R2, and F4-R4. Clusters have been marked with the subtypes they contain. Reference cluster: Stx2a with F1-R1.

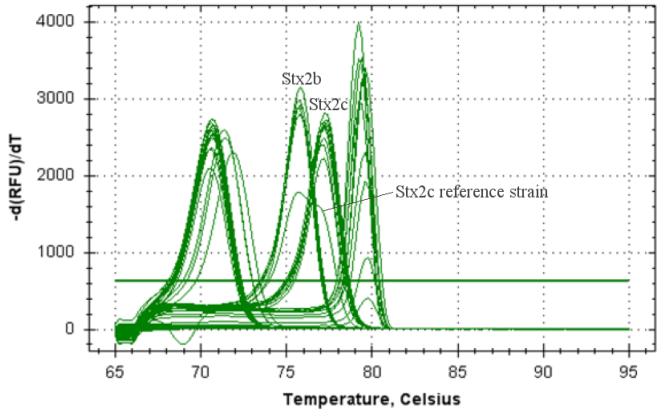


Figure 22: Melt peaks of various subtypes amplified in real-time PCR. The Stx2c reference strain have two melt peaks which overlaps with the melt peaks of Stx2b, Stx2c from pure culture samples.

With the exceptions of the samples mentioned above, all samples can be typed according to the schemes in figures 23 and 24 and table 14. Samples containing two subtypes can be distinguished from samples containing only one subtype, but it is not possible to say whether the sample contains both Stx2a and Stx2c or both Stx2a and Stx2d because some samples containing both Stx2a and Stx2d may show the same profile as samples containing both Stx2a and Stx2c.

Table 14: An overview of what type of signals each of the primer pairs produce with real-time PCR and HRMA for each subtype. + means amplification, - means no amplification, and letters indicate that there is an amplification which has a distinct and recognisable HRM-profile.

	Sample subtype							
Primer pair	Stx2a	Stx2c	Stx2d	Stx2b	Stx2e	Stx2g	Stx2a+c	Stx2a+d
F1-R1	+	+ AC	+ DE	-	+ DE	+ G	+ AC	+
F3-R1	-	+ CD	+	+	-	-	+ CD	+
F4-R4	+	-	-	-	-	-	+	+

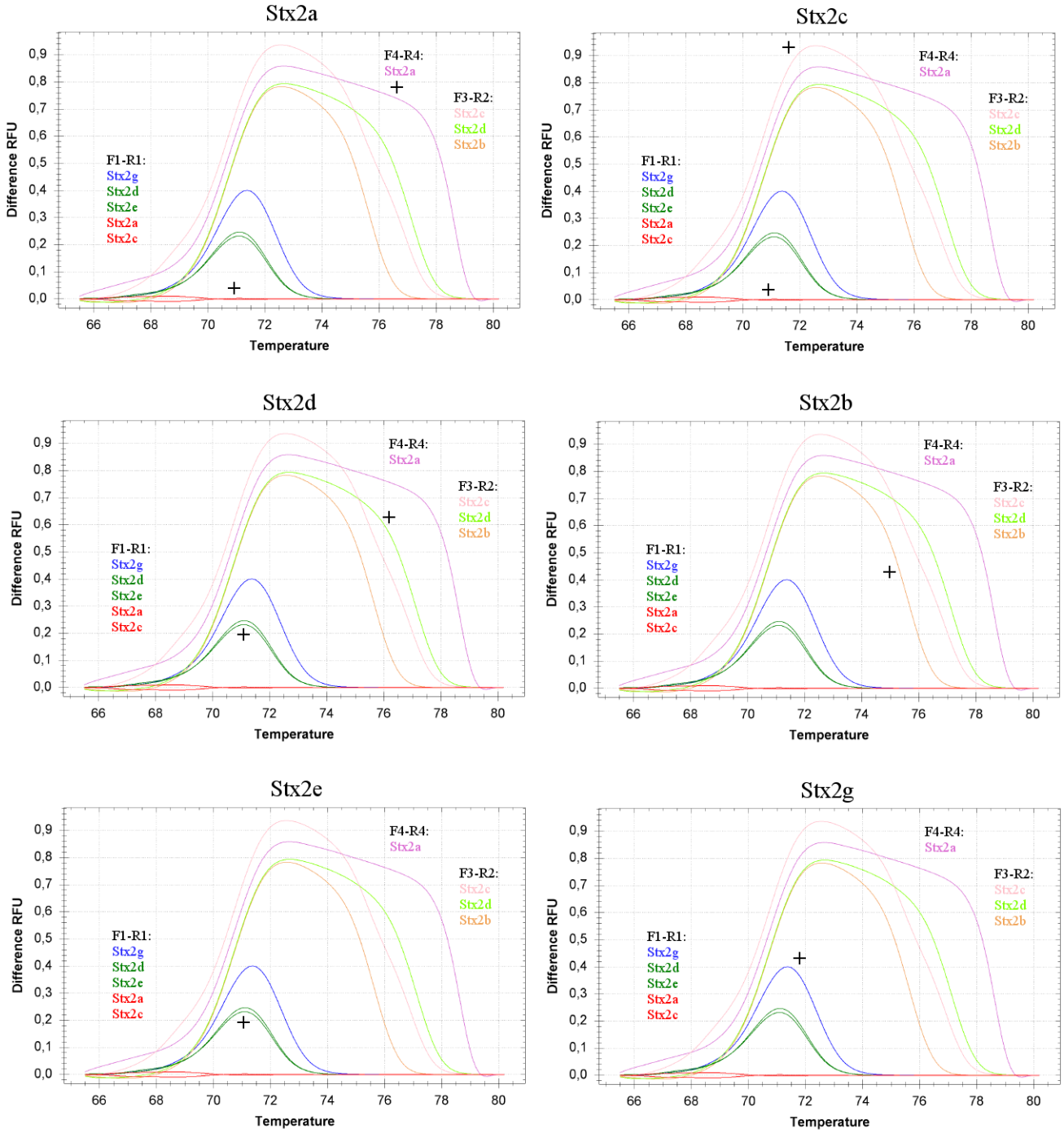


Figure 23: An overview of expected HRM profiles for each subtype after real-time PCR with primer pairs F1-R1, F3-R2, and F4-R4 along with amplified reference strains. Reference cluster: Stx2a with F1-R1. Lets say that a sample contains Stx2a (top left). The sample is run in real-time PCR with HRM in three separate PCR tubes containing primer pairs F1-R1, F3-R2, and F4-R4 respectively. The sample will be amplified with some of the primers, producing distinctive curves (marked +). These curves make a HRM profile which can be used to determine the subtype.

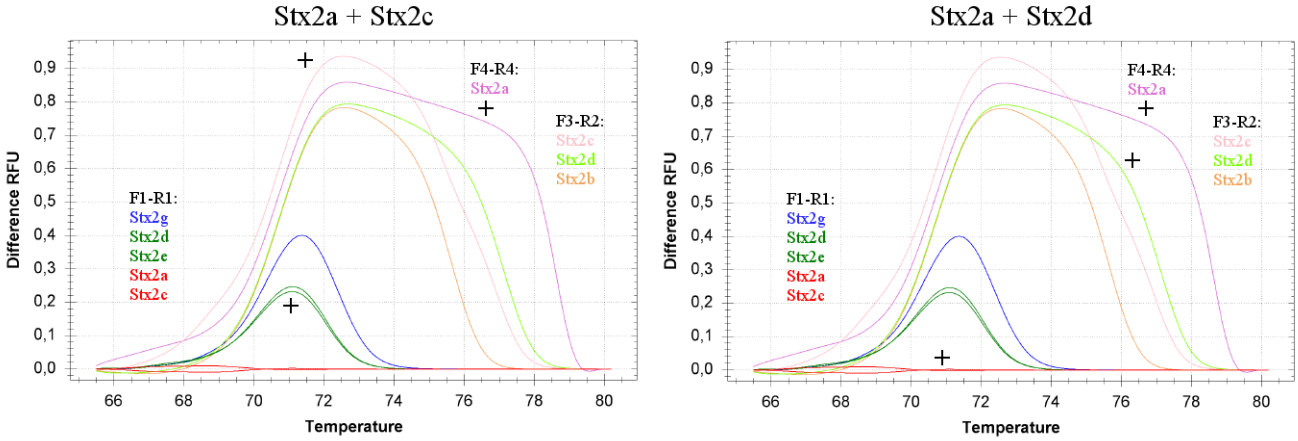


Figure 24: An overview of expected HRM profiles for samples containing more than one subtype after real-time PCR with primer pairs F1-R1, F3-R2, and F4-R4 along with amplified reference strains. Reference cluster: Stx2a with F1-R1.

5 Discussion

The goal of this project was to find an effective way of differentiating between Stx2 subtypes using real-time PCR and HRMA. Two different sets of primer pairs were tested and evaluated. The first set was only able to distinguish Stx2a, Stx2c, and Stx2d from Stx2b, Stx2e, Stx2g. Neither was it possible to say whether a sample consisted of more than just one subtype. The second set showed more promise. With some exeptions (discussed below), it could differentiate between subtypes Stx2a, Stx2c, Stx2d, Stx2b, Stx2e, Stx2g, and determine if a sample contained Stx2a and Stx2c or Sx2a and Stx2d.

In this study, the nucleotide sequences published by Scheutz *et al* were used as reference sequences since these had been characterized extensively at the WHO Collaborating Centre for Reference and Research on Escherichia and Klebsiella, Statens Serum Institut, Denmark. Initially, the Stx2 nucleotide sequences consisted of more variants than those listed in the appendix. They included strains that were designated Stx2 but not subtyped, in addition to strains which had been subtyped. Using strains that had not been subtyped posed the obvious problem of finding out what subtype they belonged to, and it was also desirable to verify if the strains that had been subtyped had actually been subtyped correctly. An attempt to subtype them was done by creating a phylogenetic tree with JalView using Stx2-variants from the Scheutz study [29]. The strains in question were then added to this tree to see which Stx2 clades they belonged to. Nearly all strains were placed in separate clades containing just one subtype, but a few deviated, including strains which had previously been subtyped by Scheutz *et al*. Considering Scheutz *et al* used protein sequences when subtyping and revising the Stx-nomeclature, this was not surprising. The selection of strains used in this study ended up being the same as that used by Scheutz *et al* because it was considered to be diverse and reliable data.

In order for HRMA to adequately separate the seven different Stx2 subtypes, their sequences had to be largely conserved, but contain some differences within each subtype that would make it possible to differentiate them. Ideally, one primer pair could have accomplished this, but it proved very difficult to find a suitable target region that was conserved in all the strains, and flanking regions that were uniquely conserved for each subtype. Moreover, it was impossible to find *one* region producing amplicons with SNPs uniquely conserved within a subtype. It also had to be taken into account that some samples may include more than one subtype.

First, three primer pairs, acdF1-R1, begF2-R2, and fF1-R1 were designed, which produced rather large amplicons containing multiple SNPs, neither of which is very ideal for HRMA. This was done on the basis that HRMA could make a more general distinction between the subtypes. For example: acdF1-R1 targeted Stx2a, Stx2c and Stx2d producing a 244 bp PCR product. The Stx2a strains, with their highly conserved sequences, would then produce the same HRMA profile, and it would be different from that of Stx2c and Stx2d. In conventional PCR, primer pair acdF1-R1 produced a 1000 bp product with Stx2g, about 750 bp larger than the expected products of Stx2a, Stx2c, and Stx2d. Because of the large difference in size, the melting temperature would be different enough to avoid confusing it with the products of Stx2a, Stx2c, and Stx2d in HRMA. Primer pair begF2-R2 amplified reference strains Stx2a, Stx2c, and Stx2d, in addition to the targeted subtypes, and acdF1-R1 also amplified Stx2b, see table 9. The primers were not specifically designed to avoid this. The reasoning was as follows: lets say a sample contained subtype Stx2a. It would be amplified in real-time PCR with acdF1-R1 and have the HRMA profile of Stx2a. It would also be amplified with begF2-R2, but the HRMA profile would be different than the profiles of the subtypes begF1-R1 was meant for. So, if a sample produce an Stx2a HRM profile with primer pair acdF1-R1 and

a signal which does not fit either Stx2b, Stx2e, or Stx2g with begF2-R2, a logical conclusion would then be that the sample contains Stx2a. When testing the first primer pair set with pure culture patient samples, it became clear that the primers were not able to produce distinctive HRMA profiles for their intended targets, probably due to the large amplicon size and multiple SNPs previously mentioned. The method was not suitable for telling samples containing two subtypes apart from samples containing just one subtype either. It could, however, be used to determine if a sample contained one of either Stx2a, Stx2c, stx2d, or a combinations of these, or one of either Stx2b, Stx2e or Stx2g. This was due to the different melting temperatures of the PCR products created with the primers, and would not have required HRMA.

Based on the results obtained with the first set of primer pairs, new primers were constructed: F1-R1 and F3-R2. These were designed to amplify smaller PCR products than the previous primer pairs, with carefully considered SNPs for the subtypes Stx2a, Stx2c, Stx2d, Stx2b, Stx2e, and Stx2g. In addition, an Stx2a specific primer pair was designed, F4-R4, which targeted a somewhat large region. That did not really matter because it was Stx2a specific and only made to give true or false answers (amplification or not) to whether or not Stx2a was present in a sample. These new primer pairs were not degenerate like the first set of primers. Having degenerate primers would have caused different types of SNPs in the template DNA's complementary sequences to be incorporated into the PCR product, affecting HRMA. This is illustrated in figure 25.

None of the primer pairs were designed to be able to amplify subtype Stx2f. The subtype is so different from the others that it would have required a primer pair of its own, which already exists. Since Stx2f is not an implicator of disease in humans it was not considered a critical component.

The detection limit and efficiency of the new primer pairs were calculated for the Stx2 subtypes that showed the lowest and higest Ct-values with each of the primer pairs after initial testing. In the dilution series, there was good amplification down to dilutions of $1{:}10^6$–$1{:}10^7$, but the HRM results seemed to be somewhat affected at very low concentrations of template DNA. The detection limit was therefore set to be about 6000 genomic copies per reaction, which is a $1{:}10^4$ dilution of the template DNA.

The PCR efficiency varied among the primer pairs and between subtypes. Primer pair F1-R1 with the Stx2g reference strain and F4-R4 with Stx2a (table 13) showed poor efficiency. One possible explanation for this may be that the primers have formed primer dimers, resulting in coamplification of nonspecific products. This could maybe have been detected by doing a melting analysis starting at a lower temperature, or by analysing the real-time PCR products on an agarose gel. Another explanation is pipetting errors made when creating the serial dilution. A third explanation involves contamination with inhibitory substances. In a dilution series, the DNA as well as any inhibitors present become diluted, and at higher dilutions the inhibitory effect diminishes. By excluding the most concentrated sample from the calculation, the efficiency improved, especially for primer pair F1-R1 with Stx2g. The HRM profiles of primer pair F4-R4 were not that interesting because it was only supposed to tell whether Stx2a was present in a sample. This means that alternate, degenerate primers could have been ordered without having to think about the effects it might have on HRMA. Using alternate primers for F4-R4 may improve the efficiency.

Some of the pure culture patient samples did not show the HRM profile expected from the subtype they belonged to. Sample s41 did not show an HRMA profile fitting Stx2d as expected. Looking at the subtyping scheme in table 14, it seems to fit the profile of a sample containing both Stx2a and Stx2d. Sample s21 (belonging to subtype Stx2d) showed the same profile as s41, except that there did not seem to be any amplification primer pair F4-R4. There was in fact an

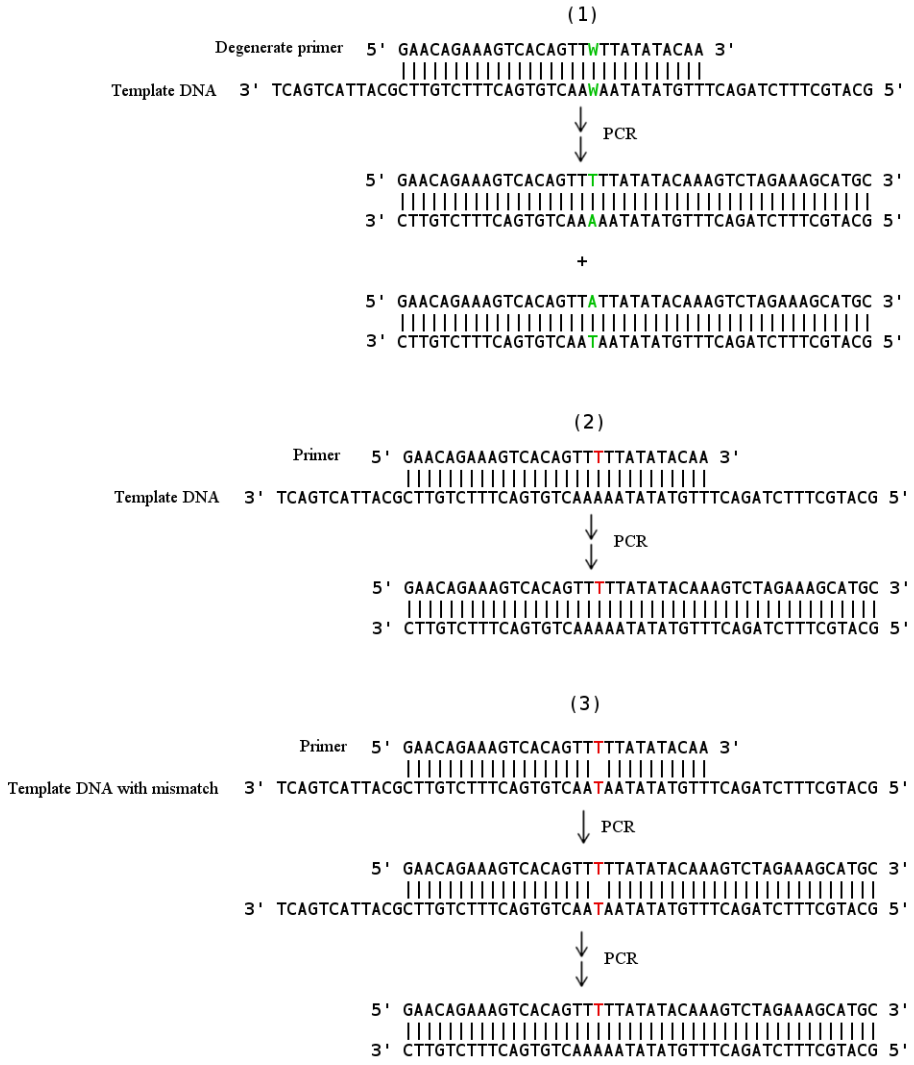


Figure 25: Having degenerate primers can affect HRMA after real-time PCR. Situation (1): a degenerate primer hybridises to two types of template DNA sequences, creating two different PCR products. Situation (2) and (3): a non-degenerate primer hybridises to the two types of template DNA, one of them containing a mismatch. In both situations, the end result after PCR will be the same PCR product.

amplification, but it was extremely weak and not included in the HRMA. The reason for why F1-R1 picked up this sample while F4-R4 did not may have been due to the poor efficiency of F4-R4. s41 and s21 may contain both subtype Stx2a and Stx2d instead of just Stx2d. Sample s44 (subtype Stx2c) showed an HRMA profile fitting that of samples containing both Stx2a and Stx2c and may contain both of these subtypes. Lastly, sample s45 (subtype Stx2e) was only amplified with the Stx2a specific primer pair F4-R4. Samples that did not show the expected HRM profile were sent to be sequenced to shed some light on these issues. Unfortunately, there was not enough time to include the results in this report, although there are indications that some of the samples may have been assigned with the wrong subtype, and it seems that s45 almost certainly does not belong to

Stx2e.

In addition to the samples just discussed, the Stx2c reference strain (accession L11079.1) deviated from the the rest of the Stx2c strains, as can be seen in figures 20, 21, and 22. Looking at figure 22, the two peaks made by the Stx2c reference strain seem to overlap with those made by the Stx2b and Stx2c strains from pure culture samples. The two melt peaks of Stx2c were also observed when doing the temperature gradient. Running the real-time PCR product on an agarose gel did did reveal anything out of the ordinary. The melt peaks were also seen when using the first set of primers. A possible explanation for this may be that the Stx2c reference strain is actually a mix of Stx2b and Stx2c.

After performing real-time PCR on the mixed culture samples, 3 out of 21 samples did not get properly amplified. One strain belonging to subtype Stx2g (accession AM904726.1) is not possible to detect with the second primer set due to a mismatch in the 3'-end. The strain is described by Scheutz *et al* [29] as being "significantly different" from the rest of Stx2g, yet it contained similarities that caused it to be designated as Stx2g. If sample s19 used in this project is the same strain as AM904726, it would explain why it was not detected by any of the primers. This sample was detectable with primers for identification of Stx2 used in routine diagnostics. Another sample, s4, was not amplified by any of the primer pairs, and neither by primers used in routine detection of Stx2. A possible explanation may be that the quality of the template DNA have been reduced over time. Sample s12 (Stx2a) showed very poor amplification with F4-R4 (probably due to the poor efficiency of F4-R4) which caused the HRMA software to mark it as a negative due to the weak signal.

Judging by the mixed culture results, it seems that the spesificity of the primers remains intact with the samples tested in this project. There were no obvious indications of amplification of any other PCR products than those produced by the *stx2* genes.

6 Conclusions and future perspectives

With the first set of primer pairs only being able to distinguish Stx2a, Stx2c, Stx2d from Stx2b, Stx2e, Stx2g, and not being able to determine whether a sample contains more than one subtype, their use is very limited and not really an option in comparison to other methods available. The second primer set showed some promise being able to differentiate between subtypes Stx2a, Stx2c, Stx2d, Stx2b, Stx2e, Stx2g, and determine that a sample contains Stx2a and Stx2c or Sx2a and Stx2d, but the issues regarding efficiency and samples showing an unexpected HRM profile must be resolved. Future work may include:

- A more thorough testing the efficiencies of all the primer pairs on each of the targeted subtypes.
- Finding a better Stx2a specific primer.
- Finding an answer as to why some of the tested samples show unexpected HRM profiles.
- Testing the primers on fecal samples.

References

1. Speirs, J. I., Stavric, S. & Konowalchuk, J. Assay of Escherichia coli heat-labile enterotoxin with vero cells. *Infection and immunity* **16,** 617–22. ISSN: 0019-9567 (Print) 0019-9567 (1977).
2. Grotiuz, G., Sirok, A., Gadea, P., Varela, G. & Schelotto, F. Shiga toxin 2-producing Acinetobacter haemolyticus associated with a case of bloody diarrhea. *Journal of clinical microbiology* **44,** 3838–41. ISSN: 0095-1137 (Print) 0095-1137 (2006).
3. Paton, A. W. & Paton, J. C. Enterobacter cloacae producing a Shiga-like toxin II-related cytotoxin associated with a case of hemolytic-uremic syndrome. *Journal of clinical microbiology* **34,** 463–5. ISSN: 0095-1137 (Print) 0095-1137 (1996).
4. Schmidt, H., Montag, M., Bockemuhl, J., Heesemann, J. & Karch, H. Shiga-like toxin II-related cytotoxins in Citrobacter freundii strains from humans and beef samples. *Infection and immunity* **61,** 534–43. ISSN: 0019-9567 (Print) 0019-9567 (1993).
5. Konowalchuk, J., Dickie, N., Stavric, S. & Speirs, J. I. Comparative studies of five heat-labile toxic products of Escherichia coli. *Infection and immunity* **22,** 644–8. ISSN: 0019-9567 (Print) 0019-9567 (1978).
6. Konowalchuk, J., Dickie, N., Stavric, S. & Speirs, J. I. Properties of an Escherichia coli cytotoxin. *Infection and immunity* **20,** 575–7. ISSN: 0019-9567 (Print) 0019-9567 (1978).
7. Konowalchuk, J., Speirs, J. I. & Stavric, S. Vero response to a cytotoxin of Escherichia coli. *Infection and immunity* **18,** 775–9. ISSN: 0019-9567 (Print) 0019-9567 (1977).
8. Pai, C. H. *et al.* Epidemiology of sporadic diarrhea due to verocytotoxin-producing Escherichia coli: a two-year prospective study. *The Journal of infectious diseases* **157,** 1054–7. ISSN: 0022-1899 (Print) 0022-1899 (1988).
9. Riley, L. W. *et al.* Hemorrhagic colitis associated with a rare Escherichia coli serotype. *The New England journal of medicine* **308,** 681–5. ISSN: 0028-4793 (Print) 0028-4793 (1983).
10. Karmali, M. A. *et al.* The association between idiopathic hemolytic uremic syndrome and infection by verotoxin-producing Escherichia coli. *The Journal of infectious diseases* **151,** 775–82. ISSN: 0022-1899 (Print) 0022-1899 (1985).
11. Karmali, M. A., Steele, B. T., Petric, M. & Lim, C. Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing Escherichia coli in stools. *Lancet* **1,** 619–20. ISSN: 0140-6736 (Print) 0140-6736 (1983).
12. Karmali, M. A., Steele, B. T., Petric, M. & Lim, C. Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing Escherichia coli in stools. *Lancet* **1,** 619–20. ISSN: 0140-6736 (Print) 0140-6736 (1983).
13. Blystad, H. *E. coli-enteritt (inkludert EHEC-infeksjon og HUS) - veileder for helsepersonell* Web Page. 2010. <`http://www.fhi.no/artikler/?id=82709`>.
14. Hussein, H. S. Prevalence and pathogenicity of Shiga toxin-producing Escherichia coli in beef cattle and their products. *Journal of animal science* **85,** E63–72. ISSN: 0021-8812 (2007).
15. Ferens, W. A. & Hovde, C. J. Escherichia coli O157:H7: animal reservoir and sources of human infection. *Foodborne pathogens and disease* **8,** 465–87. ISSN: 1535-3141 (2011).

16. O'Brien, A. D., LaVeck, G. D., Thompson, M. R. & Formal, S. B. Production of Shigella dysenteriae type 1-like cytotoxin by Escherichia coli. *The Journal of infectious diseases* **146,** 763–9. ISSN: 0022-1899 (Print) 0022-1899 (1982).

17. Smith, H. W., Green, P. & Parsell, Z. Vero cell toxins in Escherichia coli and related bacteria: transfer by phage and conjugation and toxic action in laboratory animals, chickens and pigs. *Journal of general microbiology* **129,** 3121–37. ISSN: 0022-1287 (Print) 0022-1287 (1983).

18. Strockbine, N. A. *et al.* Two toxin-converting phages from Escherichia coli O157:H7 strain 933 encode antigenically distinct toxins with similar biologic activities. *Infection and immunity* **53,** 135–40. ISSN: 0019-9567 (Print) 0019-9567 (1986).

19. Jackson, M. P., Newland, J. W., Holmes, R. K. & O'Brien, A. D. Nucleotide sequence analysis of the structural genes for Shiga-like toxin I encoded by bacteriophage 933J from Escherichia coli. *Microbial pathogenesis* **2,** 147–53. ISSN: 0882-4010 (Print) 0882-4010 (1987).

20. Asakura, H. *et al.* Detection and long-term existence of Shiga toxin (Stx)-producing Escherichia coli in sheep. *Microbiology and immunology* **42,** 683–8. ISSN: 0385-5600 (Print) 0385-5600 (1998).

21. Ohmura-Hoshino, M. *et al.* Genetic and immunological analysis of a novel variant of Shiga toxin 1 from bovine Escherichia coli strains and development of bead-ELISA to detect the variant toxin. *Microbiology and immunology* **47,** 717–25. ISSN: 0385-5600 (Print) 0385-5600 (2003).

22. Asakura, H. *et al.* Detection and genetical characterization of Shiga toxin-producing Escherichia coli from wild deer. *Microbiology and immunology* **42,** 815–22. ISSN: 0385-5600 (Print) 0385-5600 (1998).

23. Schmitt, C. K., McKee, M. L. & O'Brien, A. D. Two copies of Shiga-like toxin II-related genes common in enterohemorrhagic Escherichia coli strains are responsible for the antigenic heterogeneity of the O157:H- strain E32511. *Infection and immunity* **59,** 1065–73. ISSN: 0019-9567 (Print) 0019-9567 (1991).

24. Lin, Z. *et al.* Cloning and sequencing of two new Verotoxin 2 variant genes of Escherichia coli isolated from cases of human and bovine diarrhea. *Microbiology and immunology* **37,** 451–9. ISSN: 0385-5600 (Print) 0385-5600 (1993).

25. Weinstein, D. L., Jackson, M. P., Samuel, J. E., Holmes, R. K. & O'Brien, A. D. Cloning and sequencing of a Shiga-like toxin type II variant from Escherichia coli strain responsible for edema disease of swine. *Journal of bacteriology* **170,** 4223–30. ISSN: 0021-9193 (Print) 0021-9193 (1988).

26. Gannon, V. P., Teerling, C., Masri, S. A. & Gyles, C. L. Molecular cloning and nucleotide sequence of another variant of the Escherichia coli Shiga-like toxin II family. *Journal of general microbiology* **136,** 1125–35. ISSN: 0022-1287 (Print) 0022-1287 (1990).

27. Schmidt, H. *et al.* A new Shiga toxin 2 variant (Stx2f) from Escherichia coli isolated from pigeons. *Applied and environmental microbiology* **66,** 1205–8. ISSN: 0099-2240 (Print) 0099-2240 (2000).

28. Asakura, H. *et al.* Phylogenetic diversity and similarity of active sites of Shiga toxin (stx) in Shiga toxin-producing Escherichia coli (STEC) isolates from humans and animals. *Epidemiology and infection* **127,** 27–36. ISSN: 0950-2688 (Print) 0950-2688 (2001).

29. Scheutz, F. *et al.* Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *Journal of clinical microbiology* **50,** 2951–63. ISSN: 0095-1137 (2012).

30. Scotland, S. M., Smith, H. R., Willshaw, G. A. & Rowe, B. VERO CYTOTOXIN PRODUCTION IN STRAIN OF ESCHERICHIA COLI IS DETERMINED BY GENES CARRIED ON BACTERIOPHAGE. **322,** 216. ISSN: 0140-6736 (1983).

31. Wagner, P. L. & Waldor, M. K. Bacteriophage control of bacterial virulence. *Infection and immunity* **70,** 3985–93. ISSN: 0019-9567 (Print) 0019-9567 (2002).

32. Herold, S., Karch, H. & Schmidt, H. Shiga toxin-encoding bacteriophages–genomes in motion. *International journal of medical microbiology : IJMM* **294,** 115–21. ISSN: 1438-4221 (Print) 1438-4221 (2004).

33. Neely, M. N. & Friedman, D. I. Arrangement and functional identification of genes in the regulatory region of lambdoid phage H-19B, a carrier of a Shiga-like toxin. *Gene* **223,** 105–13. ISSN: 0378-1119 (Print) 0378-1119 (1998).

34. Plunkett, r. G., Rose, D. J., Durfee, T. J. & Blattner, F. R. Sequence of Shiga toxin 2 phage 933W from Escherichia coli O157:H7: Shiga toxin as a phage late-gene product. *Journal of bacteriology* **181,** 1767–78. ISSN: 0021-9193 (Print) 0021-9193 (1999).

35. Ethelberg, S. *et al.* Virulence factors for hemolytic uremic syndrome, Denmark. *Emerging infectious diseases* **10,** 842–7. ISSN: 1080-6040 (Print) 1080-6040 (2004).

36. Teel, L. D., Melton-Celsa, A. R., Schmitt, C. K. & O'Brien, A. D. One of two copies of the gene for the activatable shiga toxin type 2d in Escherichia coli O91:H21 strain B2F1 is associated with an inducible bacteriophage. *Infection and immunity* **70,** 4282–91. ISSN: 0019-9567 (Print) 0019-9567 (2002).

37. Hayashi, T. *et al.* Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12. *DNA research : an international journal for rapid publication of reports on genes and genomes* **8,** 11–22. ISSN: 1340-2838 (Print) 1340-2838 (2001).

38. Soborg, B. *et al.* A verocytotoxin-producing E. coli outbreak with a surprisingly high risk of haemolytic uraemic syndrome, Denmark, September-October 2012. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **18.** ISSN: 1025-496x (2013).

39. Endo, Y., Mitsui, K., Motizuki, M. & Tsurugi, K. The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and the characteristics of the modification in 28 S ribosomal RNA caused by the toxins. *The Journal of biological chemistry* **262,** 5908–12. ISSN: 0021-9258 (Print) 0021-9258 (1987).

40. Jacewicz, M., Clausen, H., Nudelman, E., Donohue-Rolfe, A. & Keusch, G. T. Pathogenesis of shigella diarrhea. XI. Isolation of a shigella toxin-binding glycolipid from rabbit jejunum and HeLa cells and its identification as globotriaosylceramide. *The Journal of experimental medicine* **163,** 1391–404. ISSN: 0022-1007 (Print) 0022-1007 (1986).

41. Endo, Y. *et al.* Site of action of a Vero toxin (VT2) from Escherichia coli O157:H7 and of Shiga toxin on eukaryotic ribosomes. **171,** 45–50. ISSN: 1432-1033 (1988).

42. Conrady, D. G. *et al.* Molecular basis of differential B-pentamer stability of Shiga toxins 1 and 2. *PloS one* **5,** e15153. ISSN: 1932-6203 (2010).

43. Fraser, M. E. *et al.* Structure of shiga toxin type 2 (Stx2) from Escherichia coli O157:H7. *The Journal of biological chemistry* **279,** 27511–7. ISSN: 0021-9258 (Print) 0021-9258 (2004).

44. Jackson, M. P., Neill, R. J., O'Brien, A. D., Holmes, R. K. & Newland, J. W. Nucleotide sequence analysis and comparison of the structural genes for Shiga-like toxin I and Shiga-like toxin II encoded by bacteriophages from Escherichia coli 933. **44,** 109–114. ISSN: 0378-1097 (1987).

45. Garmendia, J., Frankel, G. & Crepin, V. F. Enteropathogenic and enterohemorrhagic Escherichia coli infections: translocation, translocation, translocation. *Infection and immunity* **73,** 2573–85. ISSN: 0019-9567 (Print) 0019-9567 (2005).

46. Melton-Celsa, A., Mohawk, K., Teel, L. & O'Brien, A. Pathogenesis of Shiga-toxin producing escherichia coli. *Current topics in microbiology and immunology* **357,** 67–103. ISSN: 0070-217X (Print) 0070-217x (2012).

47. Tree, J. J., Wolfson, E. B., Wang, D., Roe, A. J. & Gally, D. L. Controlling injection: regulation of type III secretion in enterohaemorrhagic Escherichia coli. *Trends in microbiology* **17,** 361–70. ISSN: 0966-842x (2009).

48. Beutin, L. & Martin, A. Outbreak of Shiga Toxin Producing Escherichia coli (STEC) O104:H4 Infection in Germany Causes a Paradigm Shift with Regard to Human Pathogenicity of STEC Strains. **75,** 408–418 (2012).

49. Siegler, R. L. *et al.* Response to Shiga toxin 1 and 2 in a baboon model of hemolytic uremic syndrome. *Pediatric nephrology (Berlin, Germany)* **18,** 92–6. ISSN: 0931-041X (Print) 0931-041x (2003).

50. Fuller, C. A., Pellino, C. A., Flagler, M. J., Strasser, J. E. & Weiss, A. A. Shiga toxin subtypes display dramatic differences in potency. *Infection and immunity* **79,** 1329–37. ISSN: 0019-9567 (2011).

51. Caprioli, A., Luzzi, I., Gianviti, A., Russmann, H. & Karch, H. Pheno-genotyping of verotoxin 2 (VT2)-producing Escherichia coli causing haemorrhagic colitis and haemolytic uraemic syndrome by direct analysis of patients' stools. *Journal of medical microbiology* **43,** 348–53. ISSN: 0022-2615 (Print) 0022-2615 (1995).

52. Shaikh, N. & Tarr, P. I. Escherichia coli O157:H7 Shiga toxin-encoding bacteriophages: integrations, excisions, truncations, and evolutionary implications. *Journal of bacteriology* **185,** 3596–605. ISSN: 0021-9193 (Print) 0021-9193 (2003).

53. Friedrich, A. W. *et al.* Escherichia coli harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. *The Journal of infectious diseases* **185,** 74–84. ISSN: 0022-1899 (Print) 0022-1899 (2002).

54. Pierard, D., Muyldermans, G., Moriau, L., Stevens, D. & Lauwers, S. Identification of new verocytotoxin type 2 variant B-subunit genes in human and animal Escherichia coli isolates. *Journal of clinical microbiology* **36,** 3317–22. ISSN: 0095-1137 (Print) 0095-1137 (1998).

55. Pradel, N. *et al.* Heterogeneity of Shiga toxin-producing Escherichia coli strains isolated from hemolytic-uremic syndrome patients, cattle, and food samples in central France. *Applied and environmental microbiology* **67,** 2460–8. ISSN: 0099-2240 (Print) 0099-2240 (2001).

56. Lindgren, S. W., Melton, A. R. & O'Brien, A. D. Virulence of enterohemorrhagic Escherichia coli O91:H21 clinical isolates in an orally infected mouse model. *Infection and immunity* **61,** 3832–42. ISSN: 0019-9567 (Print) 0019-9567 (1993).

57. Ito, H., Terai, A., Kurazono, H., Takeda, Y. & Nishibuchi, M. Cloning and nucleotide sequencing of Vero toxin 2 variant genes from Escherichia coli O91:H21 isolated from a patient with the hemolytic uremic syndrome. *Microbial pathogenesis* **8,** 47–60. ISSN: 0882-4010 (Print) 0882-4010 (1990).

58. Persson, S., Olsen, K. E., Ethelberg, S. & Scheutz, F. Subtyping method for Escherichia coli shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. *Journal of clinical microbiology* **45,** 2020–4. ISSN: 0095-1137 (Print) 0095-1137 (2007).

59. Bennett, A. R., MacPhee, S. & Betts, R. P. The isolation and detection of Escherichia coli O157 by use of immunomagnetic separation and immunoassay procedures. *Letters in applied microbiology* **22,** 237–43. ISSN: 0266-8254 (Print) 0266-8254 (1996).

60. Mullis, K. *et al.* Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor symposia on quantitative biology* **51 Pt 1,** 263–73. ISSN: 0091-7451 (Print) 0091-7451 (1986).

61. Pelt-Verkuil, E., Belkum, A. & Hays, J. P. *Principles and Technical Aspects of PCR Amplification* (2008).

62. Higuchi, R., Fockler, C., Dollinger, G. & Watson, R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Bio/technology (Nature Publishing Company)* **11,** 1026–30. ISSN: 0733-222X (Print) 0733-222x (1993).

63. Laboratories, B.-R. *Real-Time PCR Applications Guide* <`http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_5279.pdf`> (2006).

64. Wittwer, C. T. *et al.* The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *BioTechniques* **22,** 176–81. ISSN: 0736-6205 (Print) 0736-6205 (1997).

65. Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G. & Pryor, R. J. High-resolution genotyping by amplicon melting analysis using LCGreen. *Clinical chemistry* **49,** 853–60. ISSN: 0009-9147 (Print) 0009-9147 (2003).

66. Gundry, C. N. *et al.* Amplicon melting analysis with labeled primers: a closed-tube method for differentiating homozygotes and heterozygotes. *Clinical chemistry* **49,** 396–406. ISSN: 0009-9147 (Print) 0009-9147 (2003).

67. Fisher, C., Meng, R., Bizouarn, F. & Scott, R. *High Resolution Melt Parameter Considerations for Optimal Data Resolution* Electronic Article. 2010. <`http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_6009.pdf`>.

68. Herrmann, M. G., Durtschi, J. D., Bromley, L. K., Wittwer, C. T. & Voelkerding, K. V. Amplicon DNA melting analysis for mutation scanning and genotyping: cross-platform comparison of instruments and dyes. *Clinical chemistry* **52,** 494–503. ISSN: 0009-9147 (Print) 0009-9147 (2006).

69. Herrmann, M. G., Durtschi, J. D., Wittwer, C. T. & Voelkerding, K. V. Expanded instrument comparison of amplicon DNA melting analysis for mutation scanning and genotyping. *Clinical chemistry* **53,** 1544–8. ISSN: 0009-9147 (Print) 0009-9147 (2007).

70. Taylor, S. *et al. A Practical Guide to High Resolution Melt Analysis Genotyping* Electronic Article. 2010.

71. *HRM - High Resoltution Melt* Web Page. 2015. <`http://hrm.gene-quantification.info/`>.

72. Stadhouders, R. *et al.* The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5' nuclease assay. *The Journal of molecular diagnostics : JMD* **12,** 109–17. ISSN: 1525-1578 (2010).

73. McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic acids research* **41,** W597–600. ISSN: 0305-1048 (2013).

A Appendix: List of accession numbers

Stx2a Accession numbers	Stx2c Accession numbers	Stx2d Accession numbers
AY443052.1	DQ235774.1	AF479828.1
EF441619.1	M59432.1	AF479829.2
EF441609.1	EF441603.1	FM998855.1
EF441618.1	AB015057.1	FN182287.1
FM998851.1	AF461167.1	X61283.1
FM998852.1	EF441604.1	AF500189.1
GQ429170.1	AB071845.1	AF500190.1
GQ429163.1	AF291819.1	EF441621.1
GQ429168.1	AY443045.1	EF584538.1
EF441613.1	L11079.1	GQ429172.1
EF441599.1	EU086525.1	FM998848.1
AB035143.1	AM982821.2	AF500188.1
AB030484.1	FM177471.1	AF500192.1
AF461171.1	AY443043.1	EF441605.1
AF524944.1	AY443049.1	AF329817.1
FM998856.1	FM998860.1	DQ059012.1
AY443054.1	AY633470.1	DQ235775.1
AJ272135.1	AY633464.1	AY443047.1
AF461173.1	AY633467.1	AF500191.1
Z37725.1	AY633469.1	AY095209.1
FM998861.1	AY633473.1	FM998840.1
FM998854.1	AY633453.1	AY633460.1
FM998839.1	AY633456.1	AY633458.1
FM998858.1		AY633457.1
FM998842.1		X67514.1 & X67515.1
EF441620.1		
GQ429162.1		
AY443057.1		
AY633459.1		
AY633471.1		
AY633472.1		
DQ344636.3		
Z50754.1		

Stx2b Accession numbers
L11078.1
X65949.1
AJ567995.1
EF441616.1
AB048222.1
AB048223.1
AJ313015.1
AJ567997.1
AB048228.1
AB048224.1
AB048229.1
AB048225.1
AB048226.1
AB048238.1
AB012102.1

Stx2e Accession numbers
AM937001.1
AB252836.1
M21534.1
AJ313016.1
AY332411.1
AY368993.1
FN182286.1
FM998846.1
X81416.1
FM998838.1
X81415.1
U72191.1
FM998844.1
X81417.1
X81418.1
AM904726.1

Stx2g Accession numbers
AY286000.1
AB048236.1
AJ966783.1
AY443060.1
AB048227.1

Stx2f Accession numbers
AB472687.1
AJ010730.1
M29153.1

B Appendix: Conserved region in Stx2a, Stx2c, and Stx2d

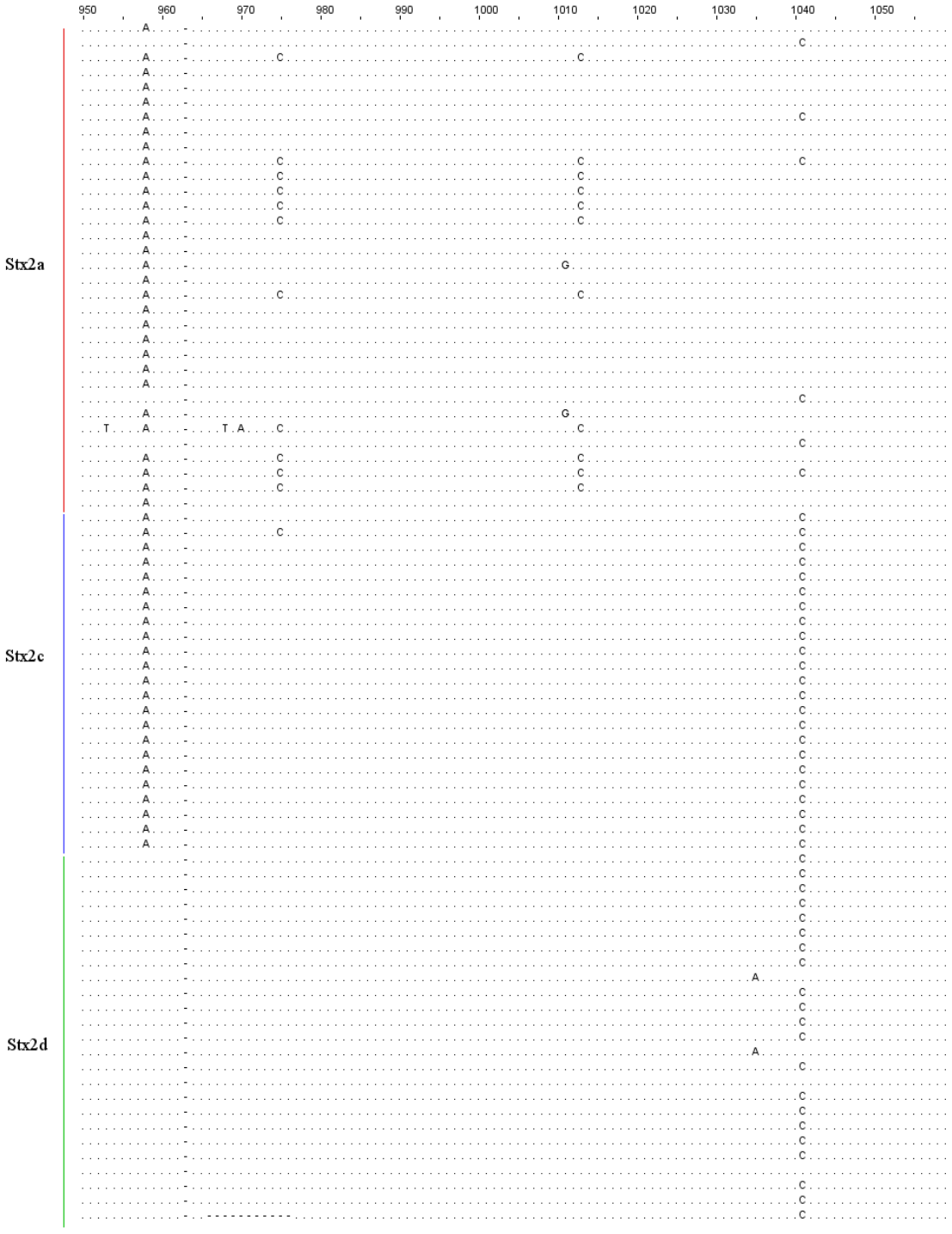


Figure 26: A highly conserverd region among the strains of subtypes Stx2a, Stx2c, and Stx2d.