

Cost-Effective TA Cloning Applied to Sanger Sequencing and HLA Allele Typing

**Anupama Cheleri NEDUVAT, Prerana Madhusudhana MURTHY,
Sudarson SUNDARRAJAN and Sriram PADMANABHAN***

Cancyte Technologies, Sri Shankara Research Center, Rangadore Memorial Hospital, Shankarapuram, Bangalore 560004, India

(* Corresponding author's e-mail: sriram.padmanabhan@cancyte.com, padsriram@gmail.com)

Received: 16 September 2017, Revised: 30 April 2018, Accepted: 26 May 2018

Abstract

Polymerase chain reaction (PCR)-based technology for clinical HLA typing involves DNA Sanger sequencing of the PCR amplified products of polymorphic loci of HLA, such as HLA-A, -B, -C, -DPB1, -DQB1, and -DRB1, in support of unrelated donor hematopoietic stem cell transplantation. The TA cloning of all the amplicons, followed by Sanger sequencing, provides a primary screening tool of potential organ donors, or to identify individuals who might have potential adverse drug responses. The Class I exon 2 and 3 genes of HLA was amplified as a single fragment and cloned into a TA cloning vector pTZ57R/T, while the Class II exon 2 and 3 amplified fragments were added together in a ligation mix with the vector pTZ57R/T. Positive clones were subjected to Sanger sequencing, and HLA alleles determined using the IMGT database. Results indicate that all the exons of the HLA genes by could be cloned by the strategies described. Furthermore, we were also successful in achieving the amplification of all the desired amplicons as a multiplex PCR, thereby reducing the cost of the modified method further by almost 60 %. We present a cost-effective TA cloning strategy for achieving accurate allele typing of HLA at 2- digit resolution. Taken together, an efficient cloning methodology with a significant lower cost for accurate HLA typing presented here is encouraging. The data suggests that it may be employed for routine cloning of variable targets in molecular biology applications.

Keywords: Allele typing, cloning, HLA typing, PCR products, pTZ57R/T

Introduction

Routines procedure in any molecular biology laboratory is the PCR amplification of a gene of interest and the cloning the amplified product into a plasmid/expression vector. Several methods have now been developed for the cloning of PCR-amplified DNA molecules. These include cohesive-end cloning through the introduction of restriction sites at the 5'-ends of PCR primers [1], blunt-end cloning [2], ligation-independent cloning [3], and TA cloning [4], etc.

TA cloning involves the cloning of any PCR amplified product using DNA polymerases (like Taq DNA polymerase) that lack 5'-3' proofreading activity and the addition of adenosine triphosphate residue to the 3' ends of the double stranded PCR product. Such PCR products are then ligated to a linearized 'T' vector with complementary 3' T overhangs on both the strands. This methodology does not require the use of restriction enzymes for cloning and, hence, is one of the simplest and most straightforward methods to clone PCR products. The cloning methodology is bidirectional, and the clones can be screened for the rightly oriented clones by Sanger sequencing or restriction mapping [5].

TA cloning kits with high cloning efficiencies are commercially available. These include TA Cloning® Kit with pCR™2.1 vector (Thermo Fisher Scientific, Waltham, MA, USA), TOPO TA cloning

kit (Invitrogen, CA, USA), T-Vectors pMD20 and pMD19 (Takara Biotech, Kusatsu, Shiga, Japan), InsTAclone PCR Cloning Kit (pTZ57R/T-Thermo Fisher Scientific, Waltham, MA, USA), TOPO TA (Invitrogen, CA, USA), and pGEM-T vector system (Promega, WI, USA). A modified version of a TA cloning vector has also been developed by Zhao *et al.* [6]. However, these kits appear expensive for employment in processing large volume samples. In this paper, we describe a TA cloning methodology that is cost-effective and also adaptable to multiple targets.

The Major Histocompatibility Complex (MHC), also known in humans as the human leukocyte antigen (HLA) region, encompasses 7.6Mb on chromosome 6p21, and encoding 252 expressed loci, including several key immune response genes. HLA genes are among the most polymorphic in the human genome, and the changes in sequence affect the specificity of antigen presentation and histocompatibility in transplantation. Six of the HLA genes (HLA-A, -B, -C, -DQA1, -DQB1, and -DRB1) are extremely polymorphic and constitute the most important set of markers for matching patients and donors for bone marrow and kidney transplantation [7].

Polymerase chain reaction (PCR)-based technology is used for clinical HLA typing [8], and DNA sequencing of the PCR amplified products of multiple loci of HLA is being used in support of unrelated donor hematopoietic stem cell transplantation [9]. HLA alleles are designated by the locus, followed by an asterisk (*), a 2-digit number corresponding to the antigen specificity, and the assigned allele number. The 2-digit resolution (e.g., A*01) of HLA typing approximates the serological antigen groupings, while the sequence specific oligonucleotide probes (SSOP) method has been used for HLA typing at 4-digit resolution (e.g., A*01:01), which can distinguish amino acid differences [10]. Since uncertainty in read alignments to highly homologous HLA allele sequences exists, accurate HLA typing at high-resolution is essential.

It is observed that the majority of the mistyped HLA alleles are accurate at the 2-digit resolution. Overall, only 5 out of the total 564 alleles at the 2-digit resolution are incorrect, corresponding to an accuracy of 99.1 %. [11]; hence, HLA typing using Sanger sequencing data can provide a primary screening tool of potential organ donors, or to identify individuals who might have potential adverse drug responses. Moreover, allele determination is generally based on sequence alignment to the IMGT/HLA database, where there is an inherent limitation.

Also, studies have shown that specific HLA alleles are associated with a number of autoimmune diseases. These include multiple sclerosis, rheumatoid arthritis, and type I diabetes [12], and infectious diseases [13]; numerous animal studies have shown that these genes are often the major contributors to disease susceptibility or resistance [14], making HLA typing clinically relevant.

Sequence-based typing (SBT) involves PCR amplification and sequencing of specific HLA exons, with both the alleles of HLA locus amplified and sequenced; hence, to determine exactly which 2 alleles were responsible for the sequence results through Sanger sequencing is difficult to predict. This limitation is, however, resolved by cloning the PCR amplicons in a TA cloning vector and selecting bacterial recombinants for both the alleles for accurate HLA typing. Traditionally, for Class I typing, most laboratories only sequence exon 2 and 3 and, for Class II typing, only sequence exon 2, although reports have also recommended the essentiality of including both exon 2 and 3 of class II. This approach has been the standard due to the functional relevance of this region, which defines the peptide groove of Class I and Class II molecules, respectively [15]. For bone marrow registries, a low to medium resolution is the minimum requirement for screening. We were, thus, motivated to develop low-cost techniques for improving resolution, and the results are described in this paper.

Sanger sequencing is the gold-standard DNA sequencing technique, used routinely in life sciences laboratories for various workflows. During Sanger sequencing, DNA polymerases copy single-stranded DNA templates by adding nucleotides to a growing chain (extension product). Because of its suitability for routine validation of cloning experiments and PCR fragments, Sanger sequencing remains a popular technique in many laboratories across the world.

The protocol described here offers comparable data quality and outperforms traditional Sanger sequencing in terms of cost-effectiveness and multiple steps and high throughput. Although the Sanger sequencing method has proven effective for HLA typing, the perceptions of it being labor-intensive, time-consuming, and expensive can now be addressed by the strategy described in this paper. Since not all

research and diagnostic laboratories have the feasibility of or access to next-generation sequencing technologies, we have been motivated to develop an efficient protocol for the genotyping of the classical HLA genes of class I and Class II by a slight tweaking of the classical TA cloning method, and the results are described in this paper.

Materials and methods

All chemicals and reagents used in this study were of analytical grade. Oligonucleotides were synthesized at BioServe Technologies Pvt. Ltd, Hyderabad, India. Standard genomic DNA sample designated-IHWG09124 was purchased from International Histocompatibility Working Group (IHWG), Seattle, WA, USA.

The blood sample from a donor SP was collected, with appropriate consent forms maintained in our laboratory. The HLA typing of the sample-SP was determined using the PacBio system at MedGenome, Bangalore, India (**Table 1**). In-house, the genomic DNA of this sample was prepared using QIAamp columns (Qiagen, Germantown, MD, USA), since they are reported to have superior binding capability [16]; hence, we preferred to use QIAamp columns for extraction of the gDNA for this study.

All the PCRs for Class I genes of sample SP were performed using 1 x ammonium sulphate buffer, pH 8.3, 0.2 mM dNTP mix, MgCl₂ (2.5 mM), DMSO, Tween-20, and BSA at 5, 2, and 0.32 % final concentrations, respectively. The PCR conditions included an initial denaturation at 95 °C for 5 min, followed by 28 cycles of 95 °C for 30 s, 61.4 °C for 45 s, and 72 °C for 45 s, while for Class II, the extension time was 1.5 min, with other PCR parameters the same. Suitable amounts of the amplified products were then analyzed on 2 % agarose gel. The PCR of the entire HLA panel comprised the following genes. HLA-A, exon-2 (375 bp); HLA-A-exon3 (327 bp); HLA-B, exon-2 (378 bp); HLA-B-exon3 (371 bp); HLA-C, exon-2; HLA-C-exon3 (375 bp); HLA-DPB1, exon-3 (368 bp); HLA-DQB1-exon3 (350 bp); HLA-DRB1, exon-3 (351 bp); HLA-DPB1-exon2 (426 bp); HLA-DQB1, exon-2 (324 bp), and HLA-DRB1-exon 2 (307 bp) were based on the primers designed in the intronic regions, as per Itoh *et al.* [17] and Lange *et al.* [18].

For PCR screening of the TA clones, M13 vector primers was used. The M13 PCR was carried out in a total volume of 25 µl using 1 x ammonium sulphate buffer, pH 8.3 with 2.5 mM magnesium chloride, 5 % DMSO, 0.32 % BSA, 2 % Tween-20, and 0.4 µM of forward (5' GTAAAACGACGGCCAGT 3') and reverse (5' AACAGCTATGACCATG 3') primers. The PCR was carried out for 35 cycles, with an initial denaturation of 5 min at 95 °C, and 35 cycles of denaturation at 95 °C for 30 s, annealing at 60 °C for 30 s, and extension at 72 °C for 45 s. After a final extension of 7 min at 72 °C, suitable aliquots of the PCR amplicons were examined on 2 % agarose gel stained with ethidium bromide.

We adopted 2 strategies for cloning Class I and II genes. This was mainly due to close proximity of the polymorphic class I genes HLA-A, -B, and -C, and significant distant localization of the Class II genes DPB1, DQB1, and DRB1.

PCR amplification of Class I HLA genes HLA-A, -B, -C genes and TA cloning

Class I genes of HLA: A, B, and C were amplified using forward primer for the exon 2 and reverse primer for the exon 3 region, in order to get a single product of ~800 to 900 bp that included both the exon 2 and 3 regions.

The gDNA of the sample- SP was quantified using NanoDrop Spectrophotometry (Thermo Fisher Scientific, Waltham, MA, USA); 40 ng gDNA was taken and, using 0.4 µM primers of HLA-A- exon 2 forward primer (**Table 1**) and HLA-A-Exon 3 reverse primers, all the genes of Class I was amplified. For PCR of IHWG09194 sample with known HLA typing (**Table 1**), 40 ng of the purified genomic DNA was taken, and PCR was carried for the exon 2 gene of HLA-B using HLA-B gene specific primers [17] at 0.4 µM final concentrations. The PCR reaction mix remained the same as mentioned above. The PCR conditions included an initial denaturation of 95 °C for 5 min, followed by 28 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 45 s, with a final extension at 72 °C for 10 min. Suitable amounts of the amplified products were then visualized on ethidium bromide stained 2 % agarose gel.

All the 3 PCR products of HLA-A, -B, and -C genes were loaded on 2 % agarose gel separately, and gel purified (Thermo Fisher Scientific, Waltham, MA, USA) and ligated to pTZ57R/T vector with an insert: vector ratio of 1:3, as suggested by the manufacturer (Thermo Fisher Scientific, Waltham, MA, USA). The ligated products were then introduced into *E. coli* competent cells (Thermo Fisher, Waltham Scientific, MA, USA), and positive clones were selected by colony PCR using gene specific primers.

Xiao-Qing *et al.* [19] describe allele typing of HLA-B using TA cloning method and also mentioned that the probability of obtaining both kinds of allele clone products by this methodology is variable. When 6 positive clones were picked in the study, the probability of getting both the alleles was close to 96.88 %. Hence, we restricted the selection of at least 6 clones from each of the HLA sets for DNA sequencing, so that the cost of the sequencing reactions was controlled and the possibility of obtaining both the arms of the alleles was achieved.

PCR amplification of class II HLA genes and TA cloning

Class II genes of HLA, namely DPB1, DQB1, and DRB1, were PCR amplified from sample SP using exon 2 and exon 3 region specific primers, to get amplicons of ~300-400 bp length. All the 6 PCR products of Class II genes, namely, DPB1, exons 2 and 3; DQB1, exon 2 and 3, and DRB1, exons 2 and 3, were purified and ligated to the vector pTZ57R/T at the ratio of inset: vector of 3:1 using the TA cloning kit (Thermo Fisher Scientific, Waltham, MA, USA).

A minimum of twenty colonies from each set were screened by M13 PCR, followed by gene specific PCRs, and a suitable number of positive clones were DNA sequenced by Sanger sequencing.

Extraction of PCR amplified products directly from excised agarose gel for Sanger sequencing and direct PCR amplicon sequencing

To verify if one could do away with plasmid isolation prior to DNA sequencing, we tested direct elution of the amplicons of HLA-B- exon2 of IHWG09124 from agarose gels and, in the second approach, we sent the same amplicons of the HLA-B- exon2 of IHWG09124 directly for Sanger sequencing without any purification.

We tested 6 independent PCR amplicons of HLA-B-exon 2 of IHWG09124 and eluted the amplicons directly from the agarose gel using nuclease free water. This was subjected to Sanger sequencing, and the sequences obtained were submitted to the IMGT database for allele determination. Also, in another experiment, we subjected 3 aliquots of the HLA-B, exon 2 PCR amplicons for Sanger sequencing directly using the gene specific primers. In all the cases, the allele type exactly matched with the type mentioned in the IHWG database for the IHWG09124 sample (**Table 1**).

Multiplex PCR

Currently, the HLA PCR panel described in this paper comprises of 3 PCRs of Class I genes and 6 PCRs of Class II genes, to cover all the target genes, namely, HLA-A, -B, and -C, and DPB1, DQB1, and DRB1. Use of exons 2 and 3 of HLA-A, -B, and -C, and exons of Class II genes, namely, DPB1, DQB1, and DRB1, is reported to be sufficient for HLA typing [20]. In the light of these reports, we attempted to carry out a multiplex PCR for the Class I genes for exons 2 and 3, and a multiplex PCR for only the exon 2 of the class II genes DPB1, DQB1, and DRB1.

For the multiplex PCRs, the PCR reaction mix was first tested by gradient PCR from 50 to 65 °C, keeping all other conditions the same, as done for uniplex PCRs (data not shown). The optimal T_m for the multiplex PCR of Class I was found to be 53.3 °C, while for Class II exon 2, it was 61.4 °C. The PCRs were carried out as described before, and the products were examined on 2.5 % agarose gels stained with ethidium bromide.

Allele determination

The IMGT/HLA database (<http://www.ebi.ac.uk/imgt/hla>) for the allelic sequences of the genes in the HLA system provides the tools for submitting both new and confirmatory sequences to the WHO HLA Nomenclature Committee [21]. All the sequencing data from all the TA clones generated in this

study were subjected to IMGT database for designating the HLA alleles. The genotypes of each allele were obtained directly according to 2 main indexes: Score and E-value.

Results

PCR amplification of Class I HLA genes HLA-A, -B, -C genes, TA cloning, and allele determination

The HLA PCR panel, comprising of HLA-A, exon 2 and 3; HLA-B, exons 2 and 3; HLA-C- exons 2 and 3, are shown in lanes 1 to 6 in the same order. Lanes 7, 8 and 9 show PCR of exons 3 of DPB1, DQB1, and DRB1, respectively, while lanes 10, 11, and 12 represent exons 2 of DPB1, DQB1, and DRB1, respectively (**Figure 1**).

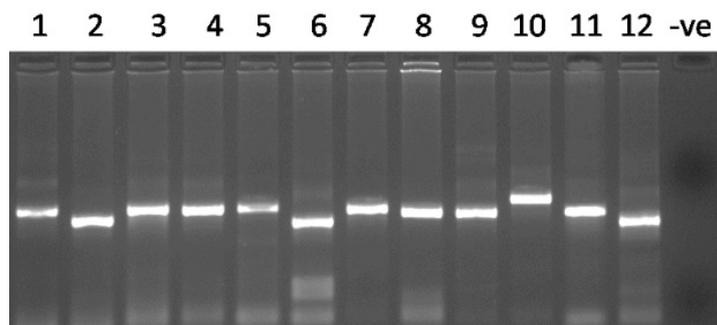


Figure 1 HLA PCR of Class I and II genes.

Lanes 1, 3, and 5 represent PCR of class I exons 2 of HLA-A, -B, and -C, while lanes 2, 4, and 5 represent the PCR of class I exons 3 of HLA-A, -B, and -C. DPB1, DQB1, and DRB1 exon 3 PCRs are shown in lanes 7, 8, and 9, while lanes 10, 11, and 12 show PCR of DPB1, BQBI, and DRB1. -ve represents no template control.

Table 1 HLA allele typing of 2 samples used in this study.

Sample	Gene	HLA Allele type
SP	HLA-A	A*23, A*11
	HLA-B	B* 58, B*27
	HLA-C	C* 03, C*15
	HLA-DPB1	DPB1 *05, DPB1 *02
	HLA-DQB1	DQB1 *03, DQB1 *06
	HLA-DRB1	DRB1 *08, DRB1 *01
IHWG09124	HLA-A	A*34, A*02
	HLA-B	B*56, B*40
	HLA-C	C*15, C*01
	HLA-DPB1	DPB1 *02, DPB1 *02
	HLA-DQB1	DQB1 *05, DQB1 *06
	HLA-DRB1	DRB1 *08, DRB1 *14

Panel A of **Figure 2** shows the PCRs of HLA-A, -B, and -C using exon 2 forward and exon 3 reverse primers, as indicated in the M & M section, while in **Figure 2**, panels B, C, and D represent the PCR screening of TA clones of HLA-A, -B, and -C using gene specific primers.

From 7 randomly chosen TA clones from each set, almost 6 of the clones showed strong PCR signals for HLA-A, -B, and -C, rendering a cloning efficiency of > 90 % in all the cases. The results are shown in **Figure 2**, where panel B represents TA clones of HLA-A-E2+E3, and panel C and D show HLA-B-E2+E3 and HLA-C-E2+E3, respectively. These clones were subjected to plasmid DNA preparation, followed by Sanger sequencing.

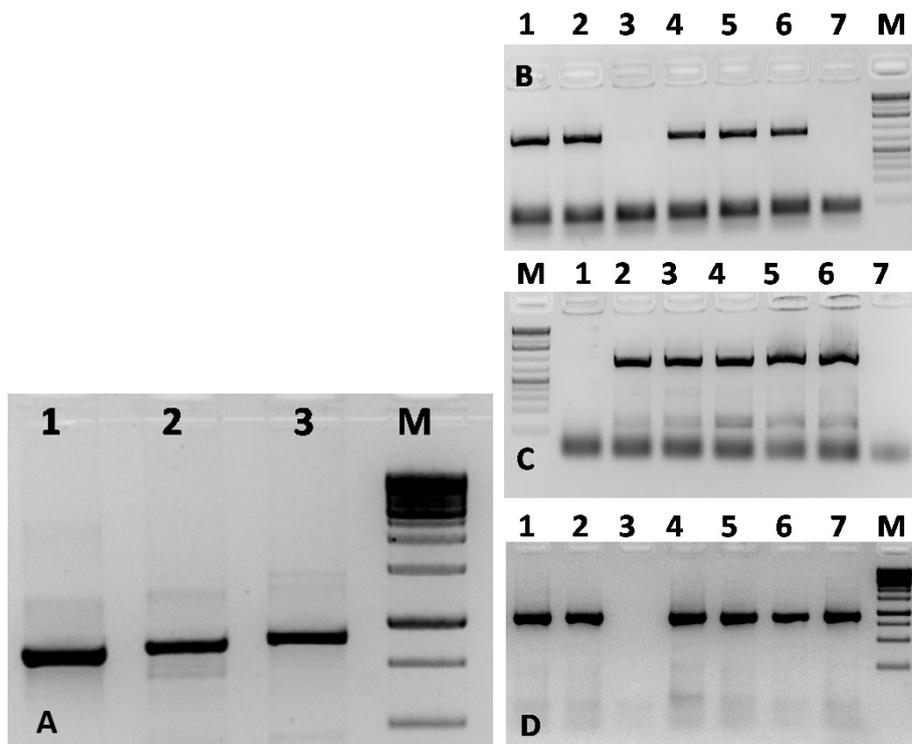


Figure 2 HLA Class I PCRs with Exon 2 forward and exon 3 reverse primers.

The sizes of the PCR products were 807 bp, 856 bp, and 892 bp, as shown in panel A, lanes 1, 2, and 3, respectively. Panels B, C, and D show the agarose gel electrophoresis of TA clones of HLA-A 2+3; HLA-B 2+3, and HLA-C 2+3. Lanes 1 to 7 show the colony PCRs, as detailed out in the M & M section. The lane M refers to DNA molecular weight marker.

Figure 3 shows the PCR screening of TA clones of Class II genes using M13 primers. Panel A, B, and C represent the PCR of twenty randomly chosen bacterial recombinants of DPB1/Exon2 and DPB1/Exon3, DQB1/Exon2 and DQB1/Exon3, DRB1/Exon2 and DRB1/Exon3, respectively.

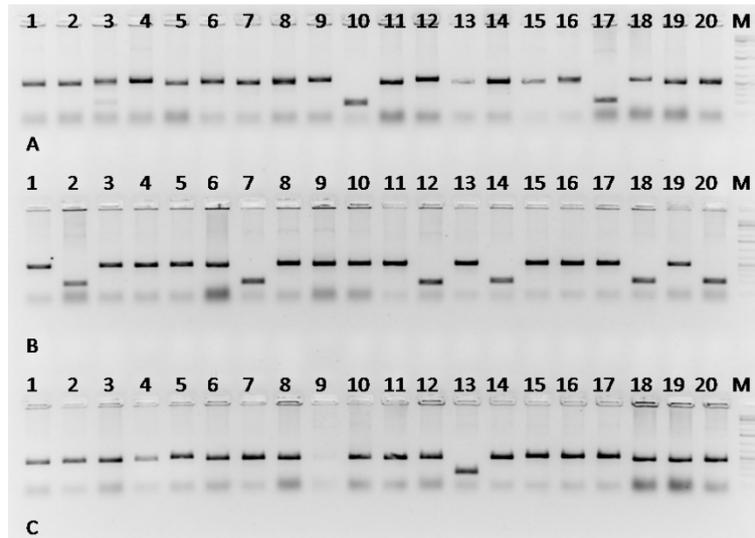


Figure 3 PCR screening of TA clones of DPB1, DQB1, and DRB1 exons 2 and 3, obtained from a single ligation mix using M13 vector primers.

Panel A shows TA clones of DPB1, panel B shows TA clones of DQB1, and panel C represents TA clones of DRB1. Note that the cloning efficiency in all the cases was more than 95 %, as evident from the figures.

It is clear from **Figure 4** that PCR of the above TA clones with DPB1 gene specific primers showed colonies # 3, 4, 9, 12, 16, 18, 19, and 20 as clones of DPB1, exon 2 (**Figure 4**, panel A), while colonies #1, 2, 5, 6, 7, 8, 11, 13, 14, 15, and 16 were clones of DPB1, exon3 (**Figure 4**, panel B). Similarly, for DQB1 exon 2, bacterial colonies that showed positive PCR signals were #6, 8, 9, and 19 (**Figure 4**, panel C), while for exon 3, the PCR positive colonies were #4, 5, 9,10, 11, 13, 15, 16, 17 (**Figure 4**, panel D).

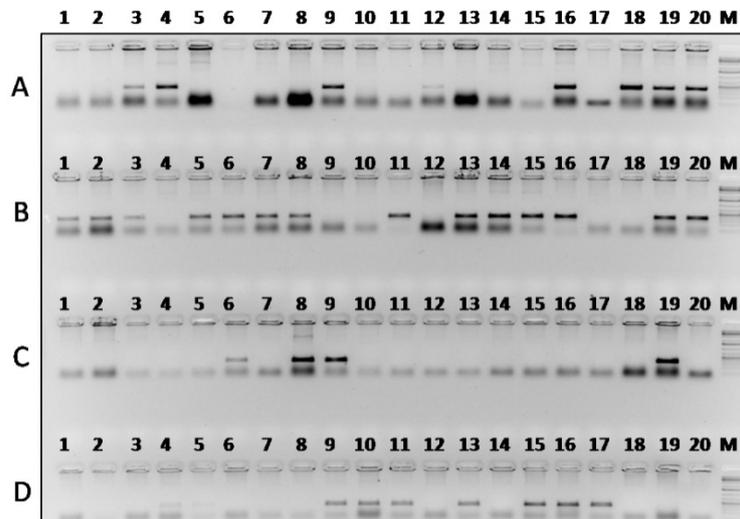


Figure 4 Gene specific PCR of TA clones of DPB1 and DQB1.

Panels A and B show PCR of exons 2 and 3 of DPB1, while panels C and D show PCRs of exons 2 and 3 of DQB1 clones. Note that one can easily see that the clones showing PCR signals in panel A do not show PCR signals in the next panel, indicating the ease in differentiating the exon 2 and 3 clones of each set. The same results were observed for DQB1 exon 2 and 3 clones.

Figure 5 shows that the bacterial colonies that were positive for DRB1, exon 2, were 15 out of 20, and that the colonies that were positive for DRB1, exon 3 were only 3 of 20, taken for PCR screening. Hence, only 6 clones of DRB1, exon 2 and 3 together were sent for DNA Sanger sequencing.

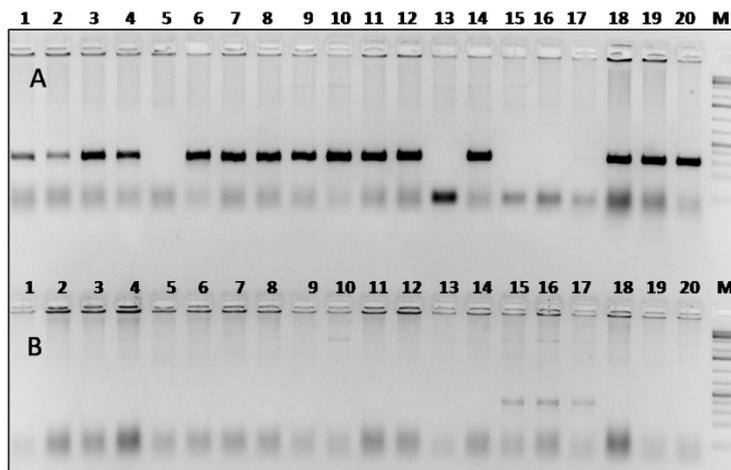


Figure 5 PCR screening of TA clones of DRB1.

Table 2 HLA allele typing of TA clones of HLA-A-E2+E3, HLA-B-E2+E3, HLA-C-E2+E3.

Bacterial TA clone no.	Gene and allele type from IMGT	Length (bp)	Score (Bits)	Identities %
HLA- A Exon 2 + Exon 3				
1	A*23	807	549.6	100
2	A*11	807	549.6	100
3	A*23	807	549.6	100
4	A*11	807	375.2	92.3
5	A*23	807	549.6	100
6	A*23	807	549.6	100
HLA-B Exon 2 + Exon 3				
1	B*58	856	541.7	99.6
2	B*58	856	549.6	100
3	B*27	856	531.8	99.6
4	B*58	856	549.6	100
5	B*27	856	541.7	100
6	B*58	856	541.7	99.6
HLA-C Exon 2 + Exon 3				
1	C*03	892	537.7	100
2	C*03	892	539.7	100
3	C*03	892	531.8	99.6
4	C*15	892	537.7	100
5	C*15	892	537.7	100
6	C*15	892	529.8	99.6

Lanes 1 to 20 of panel A show clones of exons 2 of DRB1, while panel B shows the clones of exon 3 of DRB1. Clones #6, 7, and 8 of DRB1, exon 2 and clones #15, 16, and 17 of DRB1, exon 3 were taken for Sanger sequencing.

These data clearly indicate that one could ligate 2 inserts successfully into a TA cloning vector with an appreciable cloning efficiency. Six of the PCR positive clones from all the sets were subjected to plasmid DNA preparation and DNA sequencing. The results indicated that, after the search blast with the IMGT database, accurate 2 digit (98 to 100 % identities to IMGT database) HLA type was achieved as expected, and represented in supplementary tables (**Tables 2 and 3**). From the IMGT HLA type data, one could extract information on both the allele types for a single gene (**Tables 2 and 3**) in all the cases accurately, with no ambiguity.

Extraction of PCR amplified products directly from excised agarose gel for Sanger sequencing

The PCR amplicons of HLA-B-exon 2 eluted from the agarose gel without column purification after Sanger sequencing showed the expected HLA allele typing with 98 - 100 % identity (**Table 4**). Also, the PCR amplicons that were subjected to direct Sanger sequencing showed identical HLA-B allele typing to what was achieved by using plasmid DNA for the same clones tested in this paper (**Table 4**).

Table 3 HLA allele typing of TA clones of HLA-DPB1-E2 and DPB1-E3, HLA-DQB1-E2 and DQB1-E3, and HLA-DRB1-E2 and DRB1-E3.

Bacterial TA clone no.	Gene and allele type from IMGT	Length (bp)	Score (Bits)	Identities %
HLA-DPB1 Exon 2 + Exon 3				
1	DPB1*05	426	529.8	100
2	DPB1*02	426	529.8	100
3	DPB1*02	426	513.9	99.3
4	DPB1*02	426	521.9	99.6
5	DPB1*02	368	539.7	98.9
6	DPB1*02	368	468.3	95.8
7	DPB1*02	368	555.6	99.6
8	DPB1*02	368	555.6	99.6
9	DPB1*02	368	555.6	99.6
HLA-DQB1 Exon 2 + Exon 3				
1	DQB1*06	350	521.9	99.3
2	DQB1*06	350	537.7	100
3	DQB1*06	350	557.5	99.6
4	DQB1*04	350	232.4	86.1
5	DQB1*04	350	557.5	99.6
6	DQB1*06	350	557.5	99.6
7	DQB1*04	350	565.5	100
8	DQB1*06	350	557.5	99.6
9	DQB1*06	350	557.5	99.6
10	DQB1*04	350	557.5	99.6
HLA-DRB1 Exon 2+ Exon 3				
1	DRB1*13	307	535.7	100
2	DRB1*13	307	521.9	99.6
3	DRB1*08	307	569.4	100
4	DRB1*08	307	545.6	99.0
5	DRB1*08	307	553.6	99.3
6	DRB1*08	307	554	99.6

Multiplex PCR

Figure 6, panel A shows amplification of Class I genes, HLA-A, -B, and -C in a single tube by multiplex PCR, while panel B shows the amplification of exons 2 of class II genes, namely, DPB1, DQB1, and DRB1, when multiplex PCR is employed.

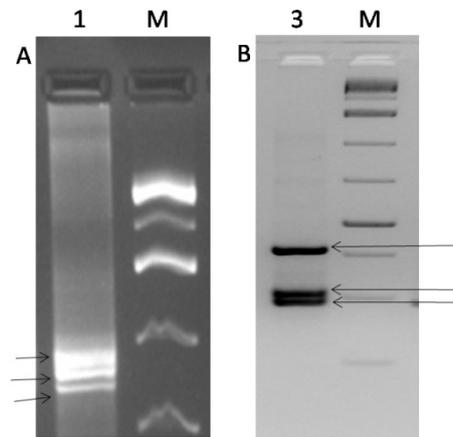


Figure 6 Multiplex PCR of Class I and Class II genes.

Panel A shows the multiplex PCR of HLA-A E2+E3, (807 bp), HLA-B-E2+E3 (856 bp), and HLA-C-E2+E3 (892 bp), while panel B represents the multiplex PCR of exon 2 of class II genes that includes HLA-DPB1, -DQB1, and -DRB1 of 426 bp, 324 bp, and 307 bp, respectively. Arrows denote the bands of interest, and M refers to DNA molecular weight marker.

Table 4 Allele typing results of HLA-B-exon 2 Sanger sequences of sample IHWG09124.

Purified plasmids				
Sample #	Gene and allele type from IMGT	Length (bp)	Score (Bits)	Identities %
1	B*56	378	531.8	99.6
2	B*56	378	539.7	100
3	B*56	378	531.8	99.6
4	B*56	378	531.8	99.6
5	B*40	378	539.7	100
6	B*56	378	523.8	99.3
Amplicons purified through gel and eluted				
Sample #	Gene and allele type from IMGT	Length (bp)	Score (Bits)	Identities %
1	B*56	378	539.7	100
2	B*56	378	531.8	99.6
3	B*40	378	539.7	100
Unpurified amplicons				
Sample #	Gene and allele type from IMGT	Length (bp)	Score (Bits)	Identities %
1	B*56	378	525.8	99.3
2	B*56	378	531.8	99.6
3	B*56	378	531.8	99.6
4	B*56	378	531.8	99.6
5	B*40	378	539.7	100
6	B*56	378	502.0	98.2

Table 5 Cost comparisons between the TA cloning for 9 target genes by original protocol and proposed protocol.

Original protocol				Proposed protocol			
Activity	Per Rxn	Cost in INR*		Activity	Per Rxn	Cost in INR*	
		Class I [#] 6 rxns	Class II [@] 3 rxns			Class I [#] 1 rxn	Class II [@] 1 rxn
PCR	44	264	132	PCR	44	44	44
Amplicon Purification	120	720	360	Amplicon Purification	-	--	--
Ligation	600	3600	1800	Ligation	600	600	600
Plasmid isolation	96	3456	1728	Plasmid isolation	-	--	--
Sanger sequencing for 6 clones of each set using a single primer [§]	200	7200	3600	Sanger sequencing for 6 clones of each set using a single primer [§]	200	3600	3600
Total	1060	15240	7620	Total	844	4244	4244
		22860				8488	

*INR- Indian rupees

[#]Using genes of HLA-A-exons 2 and 3, HLA-B- exons 2 and 3, and HLA-C exons 2 and 3

[@]Based on literature reports of using only the exon 2 of Class II genes DPB1, DQB1, and DRB1

[§]Cost calculated for carrying out unidirectional Sanger sequencing of the amplicons

Cost comparison

We have compared the cost of carrying out TA cloning of all the 6 target genes of Class I and 3 target genes of Class II, from the step of PCR amplification to the final step of DNA Sanger sequencing, between the original TA cloning protocol that is regularly followed against the proposed protocol described in this paper. The data (**Table 5**) clearly indicates an almost 60 % cost reduction in the protocol proposed here, which will have a significant impact when samples are processed on a large scale/ in large numbers.

Discussion

There are numerous reports on improvements of TA cloning in the literature [22]. Zhao *et al.* [6] report the construction of an improved TA cloning vector (pGEM-FT), by use of a restriction enzyme site XcmI to generate linear plasmids with single overhang deoxythymidine at both 3' ends for better cloning efficiency. Studies have shown that the cloning efficiency of TA cloning vectors vary considerably, based on the size of the insert used for ligation [23]. In our studies, too, the TA cloning vector pTZ57R/T vector showed better cloning efficiency with smaller sized inserts, like with Class II genes in comparison to Class I genes, supporting the observations of Palatinszky *et al.* [23].

Emonet *et al.* [24] have described a sequencing procedure where the method relies on ultrasonic shearing of PCR products, resulting in fragments 700 - 1,000 nt long that are further ligated to TA vector and sequenced. The complete sequence information is obtained approximately 5 d after the PCR step, depending on the sequencing procedure adopted. The NGS procedure requires a large sample size, and the process takes nearly 2 - 3 days for the MiSeq run, followed by the time required for bioinformatic analysis. Sanger sequencing, on the other hand, requires merely 2 - 3 days total turnaround time, where PCR, ligation, and transformation is carried out in one day, followed by sequencing the PCR amplicons the next day, with an added advantage of the feasibility of processing a single sample.

Sanger sequencing is a method of DNA sequencing based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in vitro DNA replication reaction. There are specific advantages of Sanger over NGS. These include the possibility of sequencing of single genes, using fewer toxic reagents, sequencing verification for site-directed mutagenesis or for the presence of cloned inserts, and achieving read-lengths of ~1,000 bp. However, the major limitation of the Sanger sequencing method for HLA typing is that it does not generate 2 separate haploid sequences. Hence, it is

not possible to resolve the individual HLA haplotype sequences in a diploid pair of chromosome 6, unless they are available as independent clones.

With the development of NGS, large amounts of sequencing data are becoming widely available. It is capable of cost-efficient, simultaneous interrogation of more than 100 genes at a time with low amounts of input DNA, no requirement of cloning before sequencing, and the capability of conducting multiple reactions in one lane, and performing sequence reading by using specific primers for each reaction is often considered to be superior to Sanger sequencing; however, one of the major limitations of NGS includes shorter read lengths when compared to Sanger sequencing, as well as higher error rates [25]. In addition, bioinformatic challenges include developing processes to store, retrieve, and process large amounts of data, reducing errors in image analyses and base-calling, and identifying and removing low-quality reads. Hence, the present work, highlighting the successful use of Sanger sequencing for HLA allele typing, assumes critical importance. Although NGS is cheaper than Sanger sequencing, it still remains an expensive technique for some small laboratories, since the single run cost is high if the sample size is small. Also, NGS data analysis is time consuming, and needs a large storage space and bioinformatics knowledge, due to the high amount of data resulting from the sequencing experiments [26].

It is predicted that only one out of 1081 pairs for HLA-A, 3 out of 4851 for HLA-B, and one out of 528 for HLA-C will fail to have the alleles identified in heterozygous individuals if exons 2 and 3 are analyzed together [27]. The amplified products of Class I, described here, includes the entire sequences of exon 2 and exon 3, encoding the hyper variable domains that form the peptide binding pocket of HLA class I molecules. These 2 exons encompass most of the variable sequences that distinguish alleles. Hence, our data on achieving accurate allele typing of class I genes, where both exons 2 and 3 are amplified as one fragment, does invite the possibility of revisiting this aspect in other areas too.

The HLA allele sequences are available in the IMGT/HLA database [20]. As of March 2017, there are currently 16,933 HLA and related alleles described by the HLA nomenclature and included in the IPD-IMGT/HLA database. With such a high number of alleles, genotyping is a very tough challenge. Low resolution (LR), referred to as generic typing, or 2-digit typing, corresponds to the identification of broad families of alleles (e.g., 02), and is the equivalent of serological typing (A2). While Gowda *et al.* [28] have described HLA allele typing using the Micro SSP DNA typing kit of One-Lambda (Thermo Fisher Scientific, Waltham, MA, USA) using exon 2 and 3 of HLA class I and exon 2 of HLA-DRB1 of class II for a 2 digit resolution for HLA typing, Boegel *et al.* [29] successfully applied seq2HLA to 50 individuals included in the HapMap project, yielding 100 % specificity and 94 % sensitivity at a p-value of 0.1 for 2-digit HLA types. It is recommended that selection of an unrelated donor is based on these first 2 sets of digits and in a second level of selection to use high resolution typing; hence, our present work on achieving 2-digit resolution of HL typing by the present methodology assumes clinical significance.

In this study, we have successfully demonstrated the possibility of sequencing the PCR products by the Sanger method directly, without the need to isolate plasmid DNA, making the entire process very cost-effective. Although plasmid DNA provides certain advantages, like complete sequencing of the insert, since the sequence include universal priming sites flanking inserted DNA, while approximately 20 - 50 bases of the fragment would not be sequenced over direct sequencing when sequenced from a PCR fragment.

There are certain advantages and disadvantages of sequencing plasmid DNA of the TA clones. While the additional cost of generating plasmid DNA is irrefutable, where DNA fragments are cloned before sequencing, the resulting sequence may contain parts of the cloning vector. Recently, one-step Sanger sequencing methods, such as Ampliseq and SeqSharp, have been developed that allow rapid sequencing of target genes without cloning or prior amplification that enzymatically removes chain termination products originating from one or both of the amplification primers [30]. Direct sequencing from a PCR fragment requires that the original PCR primers be used, as with plasmid DNA. The Sanger dye-terminator method is a linear amplification requiring sufficient copy numbers off the original template.

Meltzer [31] has highlighted several advantages of direct sequencing of polymerase chain reaction (PCR) products over conventional sequencing of cloned products, with speed as the greatest asset. The

steps involved include purification of PCR amplified fragments, sub-cloning, bacterial transformation, and plasmid preparation steps, that are all eliminated when the PCR amplicons are directly sequenced. The other biggest advantage of direct PCR product sequencing is that one could avoid PCR generated mutations which are amplified upon cloning prior to sequencing. Since there is an ability to sequence thousands, or even millions, of different templates at once, thereby obtaining pooled averages of mutated or polymorphic sequences, our data on right allele typing using direct PCR product sequences will be beneficial.

The main advantage of the described approach in this article is the possibility of separation of the sequences derived from each allele of one HLA locus and, hence, its usefulness in clinical laboratories focused on allele typing of HLA, that has become an essential tool in understanding pathogenesis of various diseases [32], is undebatable; although other alternative methodologies, like the use of allele-specific sequencing primers [33] and sequence specific primers [34] are reported, they are expensive, not free from unambiguous typing issues, and not easily scalable. The TA cloning methodology is an attractive alternative, since it employs the use of common gene specific primers that is easily scalable in the NGS system without any standardization, and can be applied for smaller sample volumes with accurate 2 digit typing.

We have observed that, in case of a homozygous sample, where both the alleles are expected to show similar allele typing, we could safely rely on this described TA methodology, since one could easily differentiate both the alleles by alignment analysis of the TA clone sequences using ClustalW, ensuring no allele drop out. Our results support the observations of Xiao-Qing *et al.* [19], who show that the probability of obtaining both kinds of allele clone products is > 99.6 % when 9 recombinant clones are screened for PCR and Sanger sequencing.

Conclusions

To summarize, we have successfully demonstrated the elimination of the steps of gel purification, plasmid DNA preparation, and multiple target ligation in regularly employed TA cloning, and the cost of generating a TA clone for a single sample is 60 % less than the original protocol. This is an attractive proposition for use in large-scale sequencing projects that have urged researchers on developing technical improvements to reduce the cost and time required to obtain single-sequence resolution. Our data in the suggested alternatives to creating bacterial libraries and to Sanger sequencing is an effort in this regard and would prove useful to the research community. Although the method has been validated in only 2 samples, and establishes a proof of concept, we have tested this methodology on several samples successfully. We are in the process of implementing this protocol for larger volumes of samples by establishing a high through-put screening, which will be dealt with elsewhere.

Acknowledgements

The authors also wish to thank Dr. K.N. Sridhar and Dr. Gautham Nadig for their constant support and encouragement. Cancyte Technologies Pvt. Ltd. is fully funded by Sri Sringeri Sharada Peetam, Bangalore, India.

References

- [1] V Jung, SB Pestka and S Pestka. Cloning of polymerase chain reaction-generated DNA containing terminal restriction endonuclease recognition sites. *Meth. Enzymol.* 1993; **218**, 357-62.
- [2] GL Costa and MP Weiner. Polishing with T4 or Pfu polymerase increases the efficiency of cloning of PCR fragments. *Nucleic Acids Res.* 1994; **22**, 2423.
- [3] S Kaluz and AP Flint. Ligation-independent cloning of PCR products with primers containing non base residues. *Nucleic Acids Res.* 1994; **22**, 4845.
- [4] D Marchuk, M Drumm, A Saulino and FS Collins. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.* 1991; **19**, 1154.
- [5] MY Zhou and CE Gomez-Sanchez. Universal TA cloning. *Curr. Issues Mol. Biol.* 2000; **2**, 1-7.

- [6] Y Zhao, Z Liu, S Yu, S Wen, L Hammarstrom and H Rabbani. Construction of a high efficiency PCR products cloning T vector using pGEM-5zf (+). *Avicenna J. Med. Biotechnol.* 2009; **1**, 37-9.
- [7] JL Davies, Y Kawaguchi, ST Bennett, JB Copeman, HJ Cordell, LE Pritchard, PW Reed, SCL Gough, SC Jenkins, SM Palmer, KM Balfour, BR Rowe, M Farrall, AH Barnett, SC Bain and JA Todd. Genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 1994; **371**, 130-6.
- [8] TJ White, N Arnheim and HA Erlich. The polymerase chain reaction. *Trends Genet.* 1989; **5**, 185-9.
- [9] FW Petersdorf and JA Hansen. A comprehensive approach for typing the alleles of the HLA-B locus by automated sequencing. *Tissue Antigens* 1995; **46**, 73-85.
- [10] SGE Marsh, ED Albert, WF Bodmer, RE Bontrop, B Dupont, HA Erlich, M Fernández-Viña, DE Geraghty, R Holdsworth, C K Hurley, M Lau, KW Lee, B Mach, M Maiers, WR Mayr, CR Müller, P Parham, EW Petersdorf, T Sasazuki, JL Strominger, A Svejgaard, PI Terasaki, JM Tiercy and J Trowsdale. Nomenclature for factors of the HLA system. *Tissue Antigens* 2010; **75**, 291-455.
- [11] Y Bai, M Ni, B Cooper, Y Wei and W Fury. Inference of high resolution HLA types using genome-wide RNA or DNA sequencing reads. *BMC Genomics* 2014; **15**, 325.
- [12] H Miyadera and K Tokunaga. Associations of human leukocyte antigens with autoimmune diseases: Challenges in identifying the mechanism. *J. Human Genet.* 2015; **60**, 697-702.
- [13] JM Blackwell, SE Jamieson and D Burgner. HLA and infectious diseases. *Clin. Microbiol. Rev.* 2009; **22**, 370-85.
- [14] C Wang, S Krishnakumar, J Wilhelmy, F Babrzadeh, L Stepanyan, LF Su, D Levinsonc, MA Fernandez-Viñad, RW Davisa, MM Davise and M Mindrinosa. High-throughput, high-fidelity HLA genotyping with deep sequencing. *Proc. Natl. Acad. Sci. USA.* 2012; **109**, 8676-81.
- [15] SD Adams, KC Barracchini, D Chen, FM Robbins, L Wang, P Larsen, R Luhm and DF Stroncek. Ambiguous allele combinations in HLA Class I and Class II sequence-based typing: When precise nucleotide sequencing leads to imprecise allele identification. *J. Transl. Med.* 2004; **2**, 30.
- [16] JJ Poh and SK Gan. Comparison of customized spin-column and salt-precipitation finger-prick blood DNA extraction. *Biosci. Rep.* 2014; **34**, e00145.
- [17] Y Itoh, N Mizuki, T Shimada, F Azuma, M Itakura, K Kashiwase, E Kikkawa, JK Kulski, M Satake and H Inoko. High-throughput DNA typing of HLA-A, -B, -C, and -DRB1 loci by a PCR-SSOP-Luminex method in the Japanese population. *Immunogenetics* 2005; **57**, 717-29.
- [18] V Lange, I Böhme, J Hofmann, K Lang, J Sauter, B Schöne, P Paul, V Albrecht, JM Andreas, DM Baier, J Nething, U Ehninger, C Schwarzelt, J Pingel, G Ehninger and AH Schmidt. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. *BMC Genomics* 2014; **15**, 63.
- [19] XQ Xing, YN Chu, Z Xiang, QX Song and GH Zhou. Establishment of cloning and sequencing method for high-resolution HLA-B genotype assay. *Chin. J. Anal. Chem.* 2014; **42**, 1574-9.
- [20] J Robinson, JA Halliwell, JD Hayhurst, P Flicek, P Parham and SGE Marsh. The IPD and IMGT/HLA database: Allele variant databases. *Nucleic Acids Res.* 2015; **43**, 1-9.
- [21] J Robinson, MJ Waller, P Parham, N de Groot, R Bontrop, LJ Kennedy, P Stoehr and SGE Marsh. IMGT/HLA and IMGT/MHC: Sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res.* 2003; **31**, 311-4.
- [22] S Yao, DJ Hart and Y An. Recent advances in universal TA cloning methods for use in function studies. *Protein Eng. Des. Sel.* 2016; **29**, 551-6.
- [23] M Palatinszky, M Nikolausz, D Sváb and K Márialigeti. Preferential ligation during TA-cloning of multi-template PCR products: A factor causing bias in microbial community structure analysis. *J. Microbiol. Methods* 2011; **85**, 131-6.
- [24] SF Emonet, G Grard, NM Brisbarre, GN Moureau, S Temmam, RN Charrel and X de Lamballerie. Long PCR Product Sequencing (LoPPS): A shotgun-based approach to sequence long PCR products. *Nat. Protoc.* 2007; **2**, 340-6.
- [25] TA Yap, AD Smith, R Ferraldeschi, B Al-Lazikani, P Workman and JS de Bono. Drug discovery in advanced prostate cancer: Translating biology into therapy. *Nat. Rev. Drug Dis* 2016; **15**, 699-718.

- [26] N Cereb, P Maye, S Lee, Y Kong and SY Yang. Locus-specific amplification of HLA class I genes from genomic DNA: locus-specific sequences in the first and third introns of HLA-A, -B, and -C alleles. *Tissue Antigens* 1995; **45**, 1-11.
- [27] C Lucan, LA Pop, A Florian, V Pileczki, B Petrushev, D Dima, I Frinc, I Berindan-Neagoe, A Irimie, C Berce, IS Florian, A Bojan and C Tomuleasa. HLA genotyping using next-generation sequencing. *Rom. J. Intern. Med.* 2016; **54**, 98-104.
- [28] M Gowda, S Ambardar, N Dighe, A Manjunath, C Shankaralingu, P Hirannaiah, J Harting, S Ranade, L Jagannathan and S Krishna. Comparative analyses of low, medium and high-resolution HLA typing technologies for human populations. *J. Clin. Cell Immunol.* 2016; **7**, 1-8.
- [29] S Boegel, M Lower, M Schafer, T Bukur, JD Graaf, V Boisguerin, O Türeci, M Diken, JC Castle and U Sahin. HLA typing from RNA-Seq sequence reads. *Genome Med.* 2012; **4**, 102.
- [30] D Sengupta and B Cookson. SeqSharp: A general approach for improving cycle-sequencing that facilitates a robust one-step combined amplification and sequencing method. *J. Mol. Diagn.* 2010; **12**, 272-77.
- [31] SJ Meltzer. *Direct Sequencing of Polymerase Chain Reaction Products*. In: BA White (ed.). *Methods in Molecular Biology. PCR Protocols. Current Methods and Applications*, Humana Press, Totowa, NJ, 1993, p. 137-42.
- [32] MM Prerana, CN Anupama, C Veenadhar, S Sundarrajan and S Padmanabhan. Human saliva and dried saliva spots as source of DNA for PCR based HLA typing using a combination of Taq DNA polymerase and AccuPrimeTaq polymerase. *Walailak J. Sci. & Tech.* 2020; **17**, 113-127.
- [33] CL Perng, LF Chang, WC Chien, TD Lee and JB Chang. Effectiveness and limitations of resolving HLA class I and class II by heterozygous ambiguity resolving primers (HARPs): A modified technique of sequence-based typing (SBT). *Clin. Biochem.* 2012; **45**, 1471-8.
- [34] N Matamoros, M Moraru and C Vilches. Advancing allele group-specific amplification of the complete HLA-C gene: Isolation of novel alleles from three allele groups (C*04, C*07 and C*08). *Tissue Antigens* 2013; **82**, 280-5.