Annals of Hematology (2019) 98:1279–1291
https://doi.org/10.1007/s00277-019-03626-w

ORIGINAL ARTICLE

Multiplex STR panel for assessment of chimerism following hematopoietic stem cell transplantation (HSCT)

Wui Chuen Chia[1] · Tze Sean Khoo[2] · S Fadilah S. Abdul Wahid[3] · Noor Farisah Abdul Razak[1] · Hafiza Alauddin[1] · Raja Zahratul Azma Raja Sabudin[1] · Ainoon Othman[4] · Roshida Hassan[5] · Noor Hamidah Hussin[1]

Received: 14 October 2018 / Accepted: 28 January 2019 / Published online: 19 February 2019

Abstract

Short tandem repeat (STR) analysis is used in chimerism monitoring after allogeneic hematopoietic stem cell transplantation (HSCT) for patients with various hematologic malignancies. Commercial forensic STR kits often contain loci with huge differences in power of discrimination (PD) across populations, causing some loci to be less informative for chimerism analysis in certain populations. This study aimed to construct a new STR multiplex panel with highly informative loci for efficient chimerism analysis. Thirteen STR markers which exhibit high PD (> 0.9) in at least 80% of 50 populations globally were selected to form a new panel and used in STR analysis of 253 Malaysian subjects. Cumulative power of discrimination (CPD) and combined power of exclusion (CPE) were determined from 253 Malaysian individuals. Loci informativity was assessed and compared to the commercial AmpFLSTR Identifiler PCR Amplification kit in 14 donor–recipient pairs. The new panel had detected 202 unique alleles including five novel alleles from the 253 individuals with high CPD and CPE (> 0.99999999999999999 and > 0.999999997 respectively). All loci from the new panel in the donor–recipient pair analysis showed higher than 50% informativity, while five loci from the commercial kit demonstrated lower than 50% informativity. Four loci from the new panel ranked the highest informativity. A sequenced allelic ladder which consists of 202 unique alleles from the 253 subjects was also developed to ensure accurate allele designation. The new 13-loci STR panel, thus, could serve as an additional powerful, accurate, and highly informative panel for chimerism analysis for HSCT patients.

Keywords Short tandem repeats (STR) · Microsatellite · Hematopoietic stem cell transplantation (HSCT) · Chimerism · Polymerase chain reaction (PCR) · Multiplex PCR

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00277-019-03626-w) contains supplementary material, which is available to authorized users.

✉ Noor Hamidah Hussin
nhhamidah@gmail.com

[1] Hematology Unit, Department of Pathology, Faculty of Medicine, Universiti Kebangsaan Malaysia Medical Centre, Jalan Yaacob Latif, Bandar Tun Razak, Cheras, 56000 Kuala Lumpur, Malaysia

[2] UKM Medical Molecular Biology Institute, National University of Malaysia, Jalan Yaacob Latif, Bandar Tun Razak, 56000 Cheras, Kuala Lumpur, Malaysia

[3] Cell Therapy Center, Universiti Kebangsaan Malaysia Medical Centre, Jalan Yaacob Latif, Bandar Tun Razak, Cheras, 56000 Kuala Lumpur, Malaysia

[4] Department of Medical Science II, Faculty of Medicine and Health Science, Islamic Science University of Malaysia, Kuala Lumpur, Malaysia

[5] National Blood Centre, Jalan Tun Razak, Titiwangsa, 50400 Kuala Lumpur, Malaysia

Abbreviations

BLAST	Basic Local Alignment Search Tool
CPD	Cumulative power of discrimination
CPE	Combined power of exclusion
CV	Coefficient of variance
GD	Gene diversity
He	Expected heterozygosity
Ho	Observed heterozygosity
HSCT	Hematopoietic stem cell transplantation
HWE	Hardy–Weinberg equilibrium
MEC	Mean paternity exclusion chance
PCR	Polymerase chain reaction
PD_F	Power of discrimination for female
PD_M	Power of discrimination for male
PE	Probability of exclusion
PI	Typical paternity index
PIC	Polymorphism information content
PM_F	Matching probability for female

PM_M	Matching probability for male
SD	Standard deviation
STR	Short tandem repeat
UKMMC	Universiti Kebangsaan Malaysia Medical Centre

Introduction

Allogeneic hematopoietic stem cell transplantation (HSCT) has been an important modality to cure various hematologic diseases [1, 2]. Chimerism analysis is important post HSCT to determine engraftment of the infused donor stem cells and to monitor disease relapse [3, 4]. Several methods are available for chimerism analysis but genotyping of STRs which are repeat units of 2 to 7 bp in length is most commonly used due to their highly polymorphic nature, allowing easy differentiation between DNA of different individuals [3, 5–8].

Currently, commercial STR kits with up to 16 STR markers are available and were designed to cover diverse global human population [9, 10]. Although the frequency of alleles differs from one population to another [11, 12], the inclusion of loci with huge differences in power of discrimination (PD) across different population are suboptimal as some loci often become less informative for chimerism analysis in certain populations [13–15]. As a result, some populations are unable to benefit fully from the panel.

This paper described the development of a 13-loci STR panel and its efficiency in chimerism analysis was determined in comparison with the commercial AmpFLSTR Identifiler kit (Applied Biosystems, USA). We employ a strategy to select loci which were previously shown to have a high PD (PD > 0.9) in at least 80% of various populations to form a balanced multiplex STR panel in terms of different populations. The resulting 13-loci STR panel could potentially serve as an alternative to chimerism analysis for Malaysians and possibly other populations.

Materials and methods

Ethics statement and sample collection

The study was approved by the Ethics Committee of UKM Medical Centre (UKMMC) (Ethics Committee Ref No: UKM1.5.3.5/244/ERGS/1/2012/SKK09/UKM/01/3). Two hundred fifty-three normal human whole blood samples were collected from the National Blood Centre, Malaysia, for population study. The samples were collected from people of the three broad ethnic groups in Malaysia (145 Malays, 78 Chinese, and 30 Indians). DNA was isolated from whole blood using QIAamp DNA Blood Mini Kit (QIAGEN, Germany). For panel comparison, 14 sets of HSCT archive DNA samples were obtained from Hospital Canselor Tuanku Muhriz, UKMMC. Each set consists of donor's DNA and recipient's DNA of pre- and post-transplant.

Selection of STR loci

The selection of STR loci was based on loci which have high PD (> 0.90) in > 80% of 50 populations globally (Supplementary Table 1). D19S433, D21S11, and D7S820 were excluded from the panel due to problems in primers design and optimizations. The resulting 13-loci STR panel consists of nine autosomal STRs, three X-STRs, and one Y-STR. The inclusion of Y-STR locus enables gender determination on top of serving as STR marker.

Primer design

Primers specific for four selected loci (DXS10011, DXS10135, D2S1338, and DXS8377) were designed using Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) and NCBI-Primer BLAST (http://blast.ncbi.nlm.nih.gov/). Primers were designed so that the size of all theoretical amplicons of one locus does not overlap with amplicons of other loci with the same fluorophore. Primer pairs were also screened for intra- and inter-primers secondary structure using NetPrimer (PREMIER Biosoft, USA). The 5′ end of forward primers for each locus was labeled with either 6-FAM™, VIC®, NED™, and PET® (Applied Biosystems, USA). Primers sequences of all loci are shown in Table 1.

Multiplex PCR assay for STR analysis

The 13-loci STR panel works in two separate PCR reactions, a 12-plex and a singleplex reaction. The DXS10011 was singled out to be amplified in a separate singleplex reaction as it was found to cross-react with primers of other loci. PCR amplification was performed by using Thermo Scientific™ Phusion™ Hot Start II High-Fidelity DNA Polymerase kit (Thermo Fisher Scientific, USA) in a final volume of 20 μL. PCR master mix was prepared by mixing 4.0 μL 5× Phusion HF Buffer, 0.4 μL of 10 mM dNTPs, 0.2 μL of Phusion Hot Start II DNA Polymerase, primers, 2 ng/μL of template DNA, and PCR-grade water. The final optimum concentration of primers for all 13 STR loci is listed in Table 1. After an initial denaturation step at 98 °C for 30 s, 27-cycle amplification was performed using the following profile: denaturation at 98 °C for 8 s, annealing at 63 °C for 15 s and extension at 72 °C for 25 s, and a final extension step at 72 °C for 10 min.

PCR products from both 12-plex and singleplex reactions were then pooled together in a single tube prior to capillary electrophoresis. 0.7 μL of pooled PCR product was then mixed with 9.1 μL of Hi-Di™ Formamide (Thermo Fisher Scientific, USA) and 0.2 μL of GeneScan™ 600 LIZ® Size Standard (Thermo Fisher Scientific, USA). After

Table 1 Primer sequences, fluorescence dyes, and concentrations for the 13-plex STR assay

Number	Loci	Sequences	Concentration (μM)	References
1.	vWA	F –6-FAM-GCCCTAGTGGATGATAAGAA TAATCAGTATGTG	0.11	[16]
		R –GGACAGATGATAAATACATAGGATGG ATGG		
2.	D8S1179	F –6-FAM-ATTGCAACTTATATGTATTT TTGTATTTCATG	0.27	[16]
		R –ACCAAATTGTGTTCATGAGTATAGTT TC		
3.	D16S539	F –6-FAM-GGGGGTCTAAGAGCTTGTAA AAAG	0.40	[16]
		R –GTTTGTGTGTGCATCTGTAAGCATGT ATC		
4.	FGA	F –6-FAM-GGCTGCAGGGCATAACATTA	0.13	[16]
		R –ATTCTATGACTTTGCGCTTCAGGA		
5.	DXS10011	F –VIC-CGTGGGAGAACCGTTTGAAG	0.43	Self-designed
		R –GAGCTGAGATTGCACCATT		
6.	Penta D	F –VIC-GAAGGTCGAAGCTGAAGTG	0.54	[16]
		R –ATTAGAATTCTTTAATCTGGACACAAG		
7.	DXS10135	F –NED-AGCCAAGAACATTTTTCAGTCAT	0.18	Self-designed
		R –TGACATTTGCAGTTATGTGAACCA		
8.	D18S51	F –NED-TTCTTGAGCCCAGAAGGTTA	0.36	[16]
		R –ATTCTACCAGCAACAACACAAATAAA C		
9.	Penta E	F –NED-ATTACCAACATGAAAGGGTA CCAATA	0.60	[16]
		R –TGGGTTATTAATTGAGAAAACTCCTT ACAATTT		
10.	D13S317	F –PET-ATTACAGAAGTCTGGGATGT GGAGGA	0.09	[16]
		R –GGCAGCCCAAAAAGACAGA		
11.	DYS385	F –PET-AGCATGGGTGACAGAGCTA	0.27	[17]
		R –GCCAATTACATAGTCCTCCTTTC		
12.	D2S1338	F –PET-ATACGTTCATTTCTTCCTAGCACT	0.43	Self-designed
		R –AATTCCTACTGGCCCATAATCCA		
13.	DXS8377	F –PET-GTCAGTCAAATCCATCCC	0.60	Self-designed
		R –ATACCAGTGCTCCCAGTT		

denaturation, PCR products were injected electrokinetically for 5 s at 3 kV, separated at 15 kV at a run temperature of 60 °C using POP-4® Polymer (Thermo Fisher Scientific, USA) in an ABI PRISM 3130 Genetic Analyzer (Thermo Fisher Scientific, USA). The fragment size and abundance of PCR products were determined by GeneScan 4.0 (Thermo Fisher Scientific, USA).

Determination of assay's sensitivity using mock mixed chimeric samples

The sensitivity of the panel was determined by analyzing mocked mixed chimeric samples of different chimerism states (50%, 25%, 12.5%, 6.25%, 3.12%, and 1.56%) which were prepared by mixing DNA samples of two unrelated donors at different ratios. The final concentration of DNA template was 2.0 ng/μL. Three sets of mocked mixed chimeric samples (male-to-male, female-to-female, and male-to-female) were prepared to mimic possible gender combination of donor–recipient pairs. The three sets of mock mixed chimeric samples were tested using the 13-loci STR panel. Chimerism analyses were performed using formula from previous study [18, 19]. Linearity and standard coefficient of variance (CV) curve were constructed and the sensitivity threshold of the panel was determined.

Statistical analysis

Intra- and inter-locus peak height ratio, allele, and haplotype frequencies were determined. The observed (Ho) heterozygosity, expected heterozygosity (He), Hardy–Weinberg equilibrium (HWE) exact test, and linkage disequilibrium (LD)

were analyzed by the Arlequin 3.5 software [20]. Data from both men and women were used for the calculation of Ho and He for all autosomal loci, whereas only female samples were used for X-STR loci.

The PD was analyzed using the formula from previous study [21]. For X-STR, PD was determined separately for men and women [22]. GD was calculated exclusively for Y-STR from male samples. HWE exact test was performed for all autosomal loci for both men and women. However, only female samples were used to calculate HWE for X-STR. LD was only determined for the three X-STRs (DXS10011, DXS8377, and DXS10135) as genetic LD between two markers in the human genome would not be calculated if their physical distances are more than 10 Mb [23]. For HWE and LD test, $P < 0.05$ was considered to be statistically significant. Other forensic parameters, such as the polymorphism information content (PIC), power of exclusion (PE), mean exclusion chance (MEC) for deficiency cases (Krüger's formula), normal trios (Kishida's formula), and duos (Desmarais' formula) were determined using ChrX-STR.org online database [24].

Preparation of single-allele fragments for NGS sequencing and allelic ladder construction

In order to obtain sufficient quantity of single-allele fragments for sequencing and allelic ladder construction, each of the 202 unique alleles underwent two rounds of PCR amplifications. Firstly, each unique allele was amplified in a singleplex PCR reaction followed by agarose gel electrophoresis to separate potential heterozygous products into distinct bands. PCR band containing single allele was then excised and gel purified using QIAquick Gel Extraction Kit (QIAGEN, Germany). Purified single-allele PCR product then underwent re-amplification to obtain sufficient DNA. Non-fluorescent primers were used if the products were meant for NGS sequencing whereas fluorescent primers were used if the products were meant for allelic ladder construction. All re-amplified PCR products were purified using QIAquick PCR Purification Kit (QIAGEN, Germany) and quantified using NanoDrop 1.0 spectrophotometer (Thermo Fisher Scientific, USA).

NGS sequencing of unique alleles

For NGS sequencing, each of the 202 single-allele fragments was normalized to 20 nM, pooled together into one tube and concentrated to obtain a final concentration of > 50 ng/μL prior to sample library preparation. Sample library for sequencing was prepared according to the modified protocol of the TrueSeq DNA PCR-Free Library Preparation kit (Illumina, USA). The modified protocol excludes the DNA fragmentation, end repair, and size selection steps in the standard workflow as all PCR products were already in the range of the desired size of less than 550 bp. Briefly, 100 μL of the concentrated PCR product was mixed with 180 μL of Ampure XP reagent and the PCR product was then purified according to manufacturer's protocol. Purified PCR products were then subjected to 3′ end adenylation, adapter ligation, and library validation according to the standard library preparation workflow. The resulting sample library was diluted to 2 nM, denatured with 0.2 N fresh NaOH, diluted to 8 pM by the addition of Illumina HT1 buffer, and then mixed with 8 pM PhiX (Illumina, USA) to achieve a final PhiX abundance of 30%. The resulting library (600 μL) was loaded with read 1, read 2, and index sequencing primers on a 600-cycle (2 × 300 paired ends) reagent cartridge (Illumina, USA) and run on a MiSeq sequencer (Illumina, USA). Upon completion, the quality of sequencing reads was first assessed via FastQC tools. Sequencing reads which showed quality of less than Q20 were discarded and excluded from further analysis. The sequences of all 202 alleles were determined.

Construction of the allelic ladder

To assure correct allele designation for STR profiling, an allelic ladder was constructed based on all unique alleles detected in this study. Only unique alleles which have their STR sequences confirmed by NGS were included in the allelic ladder. To construct an allelic ladder, each of the 202 sequence-validated single-allele fragments was normalized to 5 nM, pooled into a single tube, and concentrated using a MicroVac instrument to achieve a final concentration of > 0.15 ng/μL for each allele. The resulting allelic ladder was analyzed by capillary electrophoresis to ensure all 202 alleles can be called by the instrument.

Chimerism analysis using archive HSCT samples

STR profiles from 14 donor–recipient sets of HSCT archive samples from UKMMC were determined by using both the 13-loci STR panel and the AmpFLSTR Identifiler Kit (Applied Biosystems, USA). The allelic ladder constructed in this study was run concurrently with the samples. The total number of informative loci for both kits were determined and compared.

Results

Primer design and multiplex PCR optimization

Specific and equal amplification of all loci were achieved by adjusting the concentration of primers for each locus. The primer sequences and optimum concentration of primers for all 13 STR loci in the final multiplex assay are listed in Table 1. PCR reaction was set to stop at cycle 27th as some loci in the 13-loci STR panel were found to reach plateau phase of the amplification beyond cycle 28th while all loci

were equally amplified and are yet to reach their plateau at cycle 27th (data not shown).

Accuracy and sensitivity of the 13-loci STR panel

By testing the panel against a series of mock chimeric samples with known percentage of chimerism (set 1: male-to-male, set 2: male-to-female, set 3: female-to-female), we showed that the STR panel was accurate in determining the percentage of chimerism. The number of informative loci in set 1, set 2, and set 3 which were used for calculation of chimerism was 13, 12, and 11 loci respectively. The observed chimerism percentage for each sample differed less than 1.3% from the prepared mixing ratio. A standard deviation of less than 4.4% was found for all markers. Coefficient of determination of the correlation between the three sets and theoretical value is > 0.9968 (Fig. 1a).

The sensitivity of chimerism analyses was determined by plotting the coefficient of variation against the observed chimerism percentage for each sample (Fig. 1b). CV percentage increases exponentially as the percentage of chimerism decreases. The evaluation of chimerism became less precise at low level of chimerism (< 5%); hence, the sensitivity threshold for our panel was set to 5% chimerism, which corresponds to a CV of 35%.

Statistical analysis of population data

The optimized 13-loci STR panel had shown superior robustness in STR profiling. All STR loci in all 253 unrelated individuals were successfully genotyped without any dropout. Inter- and intra-peak height ratios (PHR) of all 253 samples genotyped were within 60% threshold. Only two loci showed stutter percentage higher than 10%, which were FGA (10.6%) and DXS8377 (20.63%) (Supplementary Table 2). Allele frequencies and forensic parameters derived from the population data for all STR loci are shown in Table 2. A total of 202 unique alleles were detected across 13 STR loci. The DXS10011 had the greatest number of variants, with 43 alleles. The cumulative PDF, PDM, and PE of the 13 STR loci were 0.99999999999999991, 0.9999999999999999994, and 0.999999997 respectively.

Numerical CPD and CPE comparison of the 13-loci STR panel with the AmpFLSTR Identifiler PCR Amplification kit in other study on Malaysian population

In addition to the chimerism analysis on transplant cases, we have also compared the numerical CPD and CPE of our newly developed 13-loci STR panel with the AmpFLSTR Identifiler PCR Amplification kit reported by Seah et al. on similar

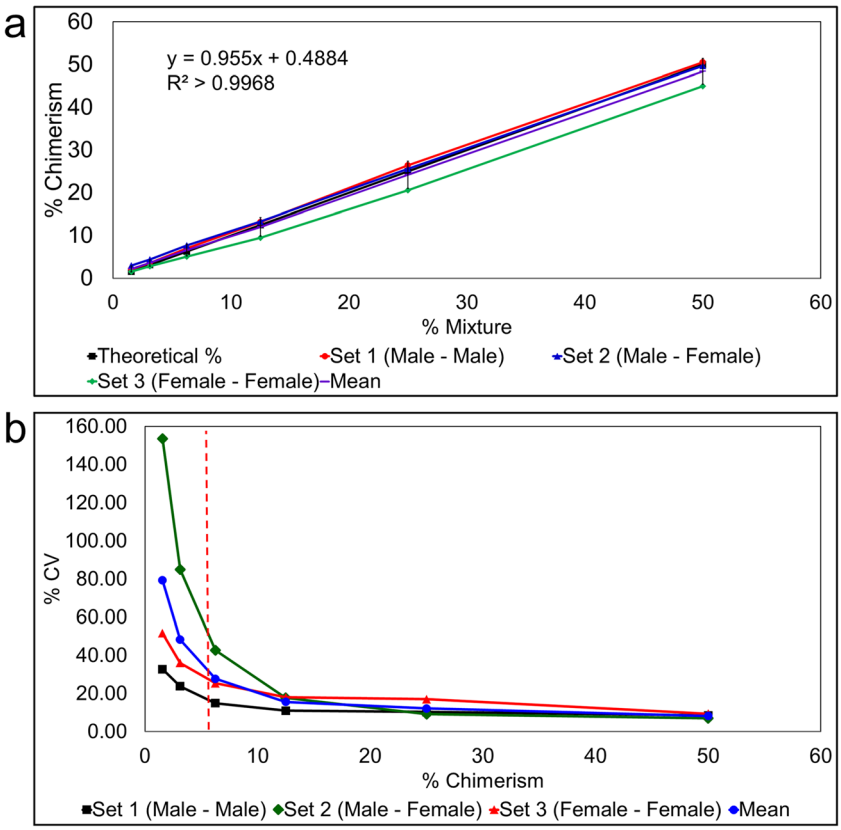

Fig. 1 **a** Percentage of chimerism obtained from each of the three sets of mock chimerism mixtures as well as the average percentage of all sets. Error bars indicate the standard deviation. The dotted line indicates the theoretically expected chimerism values of the mock chimeric mixtures. Coefficient of determination of the correlation between the three sets and theoretical value is > 0.9968. **b** Detection limit for the panel was determined by plotting percentage of chimerism over coefficient of variation for the mock chimeric samples analyzed. The vertical line at 5% indicates the detection limit for the panel

Table 2 Allele frequency distributions and forensic parameters of all STR loci in Malaysia population

Allele	vWA	D8S1179	D16S539	FGA	Penta D	D18S51	Penta E	D13S317	D2S1338	DXS10011	DXS10135	DXS8377	Haplotypes	DYS385
4	–	–	–	–	–	–	0.0494	–	–	–	–	–	–	–
6	–	–	–	–	0.0257	–	0.0099	–	–	–	–	–	–	–
7	–	–	–	–	0.0632	–	0.0020	–	–	–	–	–	–	–
8	–	0.0040	0.0119	–	0.3261	–	0.0178	0.0020	–	–	–	–	–	–
9	–	–	0.1897	–	0.1561	–	0.0217	0.2866	–	–	–	–	–	–
10	–	0.1462	0.1561	–	0.1344	0.0040	0.1621	0.1107	–	–	–	–	–	–
11	–	0.0850	0.2964	–	0.1462	0.0099	0.1344	0.1423	–	–	–	–	–	–
12	–	0.1008	0.2075	–	0.0909	0.0672	0.0711	0.2648	–	–	–	–	–	–
13	–	0.1818	0.1166	–	0.0455	0.1364	0.1206	0.1423	–	–	–	–	–	–
14	0.1937	0.1917	0.0217	–	0.0099	0.2115	0.0771	0.0435	–	–	–	–	–	–
15	0.0494	0.1818	–	–	0.0020	0.1996	0.0751	0.0079	–	–	–	–	–	–
16	0.1324	0.0771	–	–	–	0.1621	0.0810	–	0.0198	–	–	–	–	–
17	0.2787	0.0277	–	0.0020	–	0.0929	0.0632	–	0.0731	–	0.0175	–	7–16	0.0061
18	0.2312	0.0040	–	0.0178	–	0.0435	0.0553	–	0.0889	–	0.0029	–	9–16	0.0061
19	0.0909	–	–	0.0573	–	0.0336	0.0316	–	0.1917	–	0.0845	–	9–17	0.0061
20	0.0237	–	–	0.0613	–	0.0158	0.0119	–	0.1166	–	0.0671	–	11–11	0.0061
20.2	–	–	–	–	–	–	–	–	–	0.0087	–	–	11–12	0.0245
21	–	–	–	0.1462	–	0.0119	0.0059	–	0.0296	–	0.0875	–	11–14	0.0675
21.2	–	–	–	0.0079	–	–	–	–	–	–	–	–	11–15	0.0061
22	–	–	–	0.1700	–	0.0059	0.0020	–	0.0949	–	0.1166	–	11–16	0.0061
22.2	–	–	–	0.0237	–	–	–	–	–	–	–	–	–	–
23	–	–	–	0.1739	–	0.0040	0.0059	–	0.1877	–	0.0991	–	11–17	0.0061
23.2	–	–	–	0.0059	–	–	–	–	–	–	–	–	–	–
24	–	–	–	0.1364	–	0.0020	–	–	0.1146	–	0.1224	–	11–18	0.0123
24.2	–	–	–	0.0099	–	–	–	–	–	–	–	–	–	–
25	–	–	–	0.1008	–	–	–	–	0.0652	0.0058	0.0787	–	11–19	0.0123
25.2	–	–	–	0.0059	–	–	–	–	–	–	–	–	–	–
26	–	–	–	0.0573	–	–	0.0020	–	0.0178	0.0029	0.0496	–	12–12	0.0061
27	–	–	–	0.0178	–	–	–	–	–	0.0029	0.0554	–	12–13	0.0061
27.2	–	–	–	–	–	–	–	–	–	0.0029	–	–	12–14	0.0123
28	–	–	–	0.0059	–	–	–	–	–	0.0146	0.0437	–	12–16	0.0368
29	–	–	–	–	–	–	–	–	–	0.0146	0.0758	–	12–17	0.0245
29.2	–	–	–	–	–	–	–	–	–	0.0058	–	–	12–18	0.0368
30	–	–	–	–	–	–	–	–	–	0.0175	0.0321	–	12–19	0.0245
30.2	–	–	–	–	–	–	–	–	–	0.0058	–	–	12–20	0.0429
31	–	–	–	–	–	–	–	–	–	0.0262	0.0262	–	12–21	0.0123
31.2	–	–	–	–	–	–	–	–	–	0.0437	–	–	12.2–20	0.0061
32	–	–	–	–	–	–	–	–	–	0.0350	0.0204	–	13–13	0.0859
32.2	–	–	–	–	–	–	–	–	–	0.0641	–	–	13–14	0.0552
33	–	–	–	–	–	–	–	–	–	0.0350	0.0029	–	13–15	0.0123
33.2	–	–	–	–	–	–	–	–	–	0.0671	–	–	13–16	0.0307
34	–	–	–	–	–	–	–	–	–	0.0496	0.0029	–	13–17	0.0245
34.2	–	–	–	–	–	–	–	–	–	0.0321	–	–	13–18	0.0491
35	–	–	–	–	–	–	–	–	–	0.0758	0.0087	–	13–19	0.0429
35.2	–	–	–	–	–	–	–	–	–	0.0087	0.0029	–	13–20	0.0123
36	–	–	–	–	–	–	–	–	–	0.0671	–	–	13–21	0.0123
36.2	–	–	–	–	–	–	–	–	–	0.0058	–	–	14–14	0.0061

Table 2 (continued)

Allele	vWA	D8S1179	D16S539	FGA	Penta D	D18S51	Penta E	D13S317	D2S1338		DXS10011	DXS10135	DXS8377	Haplotypes	DYS385
37	–	–	–	–	–	–	–	–	–		0.0408	0.0029	–	14–15	0.0245
37.2	–	–	–	–	–	–	–	–	–		0.0058	–	–	14–17	0.0184
38	–	–	–	–	–	–	–	–	–		0.0671	–	–	14–18	0.0184
39	–	–	–	–	–	–	–	–	–		0.0408	–	–	14–19	0.0184
40	–	–	–	–	–	–	–	–	–		0.0554	–	0.0145	14–20	0.0123
41	–	–	–	–	–	–	–	–	–		0.0437	–	0.0029	15–16	0.0245
41.2	–	–	–	–	–	–	–	–	–		0.0029	–	–	15–17	0.0061
42	–	–	–	–	–	–	–	–	–		0.0350	–	0.0233	15–18	0.0368
43	–	–	–	–	–	–	–	–	–		0.0321	–	0.0640	15–19	0.0491
44	–	–	–	–	–	–	–	–	–		0.0117	–	0.0698	15–20	0.0429
45	–	–	–	–	–	–	–	–	–		0.0233	–	0.1105	15–21	0.0061
45.3	–	–	–	–	–	–	–	–	–		0.0029	–	–	15–22	0.0184
46	–	–	–	–	–	–	–	–	–		0.0087	–	0.1017	16–17	0.0061
47	–	–	–	–	–	–	–	–	–		0.0175	–	0.1221	17–19	0.0061
48	–	–	–	–	–	–	–	–	–		0.0029	–	0.1308	17–22	0.0061
48	–	–	–	–	–	–	–	–	–		0.0029	–	–	20–20	0.0061
49	–	–	–	–	–	–	–	–	–		0.0029	–	0.1105	–	–
50	–	–	–	–	–	–	–	–	–		0.0029	–	0.0901	–	–
51	–	–	–	–	–	–	–	–	–		0.0029	–	0.0610	–	–
52	–	–	–	–	–	–	–	–	–		0.0029	–	0.0436	–	–
53	–	–	–	–	–	–	–	–	–		0.0029	–	0.0320	–	–
54	–	–	–	–	–	–	–	–	–		–	–	0.0145	–	–
55	–	–	–	–	–	–	–	–	–		–	–	0.0029	–	–
56	–	–	–	–	–	–	–	–	–		–	–	0.0029	–	–
58	–	–	–	–	–	–	–	–	–		–	–	0.0029	–	–
Sum	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		1.0000	1.0000	1.0000		1.0000
PIC	0.7743	0.8335	0.7639	0.8670	0.7932	0.8373	0.8989	0.7629	0.8603		0.9518	0.9133	0.8951	0.9629 (GD)	
Ho	0.8042	0.8498	0.7905	0.8893	0.8379	0.8419	0.9210	0.7431	0.8577		0.9222	0.0810	0.0969		
He	0.8182	0.8533	0.7961	0.8807	0.8164	0.8556	0.9082	0.7946	0.8741		0.9590	0.9444	0.8901		
PE	0.6038	0.6977	0.5888	0.7526	0.6267	0.7025	0.8083	0.5862	0.7411		0.9060	0.9242	0.9081		
PI	0.0987	0.0743	0.1028	0.0605	0.0926	0.0731	0.0468	0.1035	0.0634		0.0231	0.0405	0.0484		
PD	0.9292	0.9573	0.9247	0.9704	0.9390	0.9589	0.9815	0.9307	0.9694	PD_F	0.9959	0.9877	0.9826		
PM	0.0708	0.0427	0.0753	0.0296	0.0610	0.0411	0.0185	0.0693	0.0306	PD_M	0.9513	0.9202	0.9080		
										PM_F	0.0041	0.0123	0.0174		
										PM_M	0.0487	0.0798	0.0920		
MEC1	0.6114	0.7001	0.5950	0.7565	0.6450	0.7084	0.8111	0.5955	0.7450		0.9066	0.8358	0.8039		
MEC2	0.7743	0.8335	0.7639	0.8670	0.7932	0.8373	0.8989	0.7629	0.8602		0.9518	0.9133	0.8951		
MEC3	0.7743	0.8335	0.7639	0.8670	0.7932	0.8373	0.8989	0.7629	0.8603		0.9518	0.9133	0.8951		
MEC4	0.6496	0.7272	0.6362	0.7754	0.6747	0.7332	0.8234	0.6354	0.7654		0.9102	0.8457	0.8172		
HWE	0.1947	0.4531	0.5358	0.2888	0.1196	0.2046	0.7358	0.2067	0.6899		0.0970	0.9187	0.2858		
SD	0.0004	0.0004	0.0004	0.0003	0.0002	0.0003	0.0003	0.0004	0.0003		0.0001	0.0003	0.0003		

PIC polymorphism information content, *Ho* observed heterozygosity, *He* expected heterozygosity, *PE* probability of exclusion, *PI* typical paternity index, PD_F power of discrimination for female, PD_M power of discrimination for male, PM_F matching probability for female, PM_M matching probability for male, *MEC* mean paternity exclusion chance, *MEC1* MEC Krüger, *MEC2* MEC Kishida, *MEC3* MEC Desmarais, *MEC4* MEC Desmarais Duo, *P* probability values of exact tests for Hardy–Weinberg equilibrium, *S.D.* standard deviation, *GD* gene diversity

ethnics in Malaysian population [25]. Data comparison of all three ethnics is shown in Table 3. CPD and CPE of the newly developed 13-loci STR panel were higher than AmpFLSTR Identifiler PCR Amplification kit for Malaysian population except for the CPE of the Indian ethnic group. Results demonstrated that the present STR panel is more discriminatory in chimerism analysis in Malaysian population.

Characterization of STR repetitive motives from NGS sequencing data

NGS sequencing has successfully produced a total of 12,965,606 paired-end reads. FastQ files of read 1 and read were merged and sequences of all 202 STR alleles were successfully determined. STR motif sequences of all alleles were shown in Table 4. Alleles 41.2, 45.3, 51, 52, and 53 from the DXS10011 were novel variant alleles which were previously unreported.

Construction of a sequence-validated allelic ladder

When the newly constructed allelic ladder was analyzed by capillary electrophoresis, the electropherogram showed that all peaks from the 202 alleles were present (Fig. 2). However, we observed two non-specific peaks (108 bp in the blue channel and 267 bp in the green channel) in the electropherogram. These two peaks were not found in any sample's STR electropherogram nor in the NGS sequencing results which suggests that they were most probably non-STR artifacts which do not affect the accuracy of genotyping.

Chimerism analysis by the 13-loci STR panel in comparison with the AmpFLSTR Identifiler PCR Amplification kit

A total of 14 sets of transplant cases in UKMMC were used to test the STR panel equipped with the sequence-validated allelic ladder for correct allele-typing. The results generated from our 13-loci STR panel were compared to the UKMMC laboratory's result generated from the AmpFLSTR Identifiler PCR Amplification kit. The number of informative alleles for each sample detected by both panels was determined (Fig. 3a). When we sort the loci from both panels according to the degree of informativeness, the four most informative loci (DXS10011, Penta D, Penta E and DXS8377) were originated from the 13-loci STR panel, whereas the least informative loci (D3S1358, D5S818, D7S820, TPOX, and CSF1PO) were loci which were found in the AmpFLSTR Identifiler kit (Fig. 3b). In addition, as few as six loci from our new 13-loci STR panel were sufficient to obtain three informative loci in 100% of donor–recipient pairs, whereas at least eight loci from the AmpFLSTR Identifiler kit were needed to obtain three informative loci for the same set of samples (Supplementary Fig. 1). In some other population, at least 10 loci from the AmpFLSTR Identifiler kit were needed to obtain three informative loci in 99% of donor–recipient pairs [14].

Discussion

In this study, a highly polymorphic panel was constructed by combining 13 STR markers which were shown to be highly discriminating (with PD > 0.9) in at least 80% of various

Table 3 CPD and CPE comparison of the newly constructed STR panel against AmpFLSTR Identifiler kit on Malaysian population

CPD and CPE	Newly developed panel Malay ($n = 145$) Chinese ($n = 78$) Indian ($n = 30$)	AmpFLSTR Identifiler kit [25] Malay ($n = 210$) Chinese ($n = 219$) Indian ($n = 209$)
Malay		
CPD_M	0.9999999999999999744	0.999999999999999974
CPD_F	0.999999999999999998	
CPE	0.999999996	0.9999980
Chinese		
CPD_M	0.999999999999999743	0.9999999999999993
CPD_F	0.99999999999999997	
CPE	0.99999999	0.9999989
Indian		
CPD_M	0.9999999999999924	0.999999999999999964
CPD_F	0.999999999999998	
CPE	0.999999974	0.9999990

Table 4 STR alleles and repeat motif of all alleles

Alleles	STR repeat motifs
vWA	
14	N44–(TCTA)–(TCTG)–(TCTA)–(TCTG)$_4$–(TCTA)$_3$–(TCCA)–(TCTA)$_3$–N39
15–20	N44–(TCTA)–(TCTG)$_4$–(TCTA)$_{10–15}$–N39
D8S1179	
8–13	N44–(TCTA)$_{8–13}$–N131
14–18	N44–(TCTA)$_{1–2}$–(TCTG)–(TCTA)$_{12–15}$–N131
D16S539	
8–13	N184–(GATA)$_{8–13}$–N60
FGA	
17–27	N102–(TTTC)$_3$–TTTT–TTCT–(CTTT)$_{9–19}$–CTCC–(TTCC)$_2$–N156
21.2–25.2[b]	N102–(TTTC)$_3$–TTTT–TT–(CTTT)$_{14–18}$–CTCC–(TTCC)$_2$–N156
DXS10011	
20.2–36.2[b]	N66–(GAAA)–GA–(GAAA)$_4$–GAGA–(GAAA)$_{5–19}$–GAAG–GAAA–(GGAA)$_4$–(AGAA)$_3$–N56
25–49	N66–(GAAA)$_{16–40}$–GAAG–GAAA–(GGAA)$_4$–(AGAA)$_3$–N56
37.2[b]	N66–(GAAA)$_6$–GA–(GAAA)$_{22}$–GAAG–GAAA–(GGAA)$_4$–(AGAA)$_3$–N56
41.2[a]	N66–(GAAA)$_{29}$–GA–(GAAA)$_3$–GAAG–GAAA–(GGAA)$_4$–(AGAA)$_3$–N56
45.1[a]	N66–(GAAA)$_{29}$–GAAG–GAAA–(GGAA)$_4$–AGGA–A–GAAG–GAAA–(GGAA)$_4$–(AGAA)$_3$–N56
50	N66–(GAAA)$_{28}$–GGAA–GGAA–(GAAA)$_{11}$–GAAG–GAAA–(GGAA)$_4$–(AGAA)$_3$–N56
5 1–53[a]	N66–(GAAA)$_{42–44}$–GAAG–GAAA–(GGAA)$_4$–(AGAA)$_3$–N56
Penta D	
6–15	N94–(AAAGA)$_{6–15}$–N275
DXS10135	
17–24	N55–(AAGA)$_3$–GAAAG–(GAAA)$_{14–21}$–N80
25–37	N55–(AAGA)$_3$–GAAAG–(GAAA)$_{16–28}$–(GGAA)–(GAAA)$_3$–(GGAA)–(GAAA)–N80
35.2[b]	N55–(AAGA)$_3$–GAAAG–(GAAA)$_{23}$–AA–(GAAA)$_3$–(GGAA)–(GAAA)$_3$–(GGAA)–(GAAA)–N80
D18S51	
10–24	N86–(AGAA)$_{10–24}$–N172
Penta E	
4–26	N325–(AAAAG)$_{4–26}$–N34
D13S317	
8, 12, 14, 15	N104–(TATC)$_{8–15}$–N40
9, 10, 11, 13	N104–(TATC)$_{8–12}$–(AATC)–N40
DYS385	
7–22	N182–(GAAA)$_{7–22}$–N31
12.2[b]	N182–(GAAA)$_{12}$–GA–N31
D2S1338	
16–23	N70–(TGCC)$_{6–7}$–(TTCC)$_{10–16}$–N203
24–26	N70–(TGCC)$_{6–7}$–(TTCC)$_{14–17}$–(GTCC)–(TTCC)$_2$–N203
DXS8377	
40–58	N232–(AGA)$_{19–35}$–(GGA–AGA)$_{5–7}$–(AGA)$_2$–GGA–(AGA)$_{5–6}$–N120

[a] New unreported alleles from DXS10011

[b] Variant alleles from various loci

population including European, Asian, North American, South American, Oceania, and African (Supplementary Table 1). Some conventional CODIS core loci (CSF1PO, D3S1358, and TPOX) which only showed high PD in certain limited population were thus excluded from this panel. In addition to its usefulness in discriminating donor–recipient pairs within a single population, the panel is also potentially useful to discriminate donor–recipient pairs originating from different continents across the globe as HSCT is performed with the use of donors coming from different continents nowadays [26].

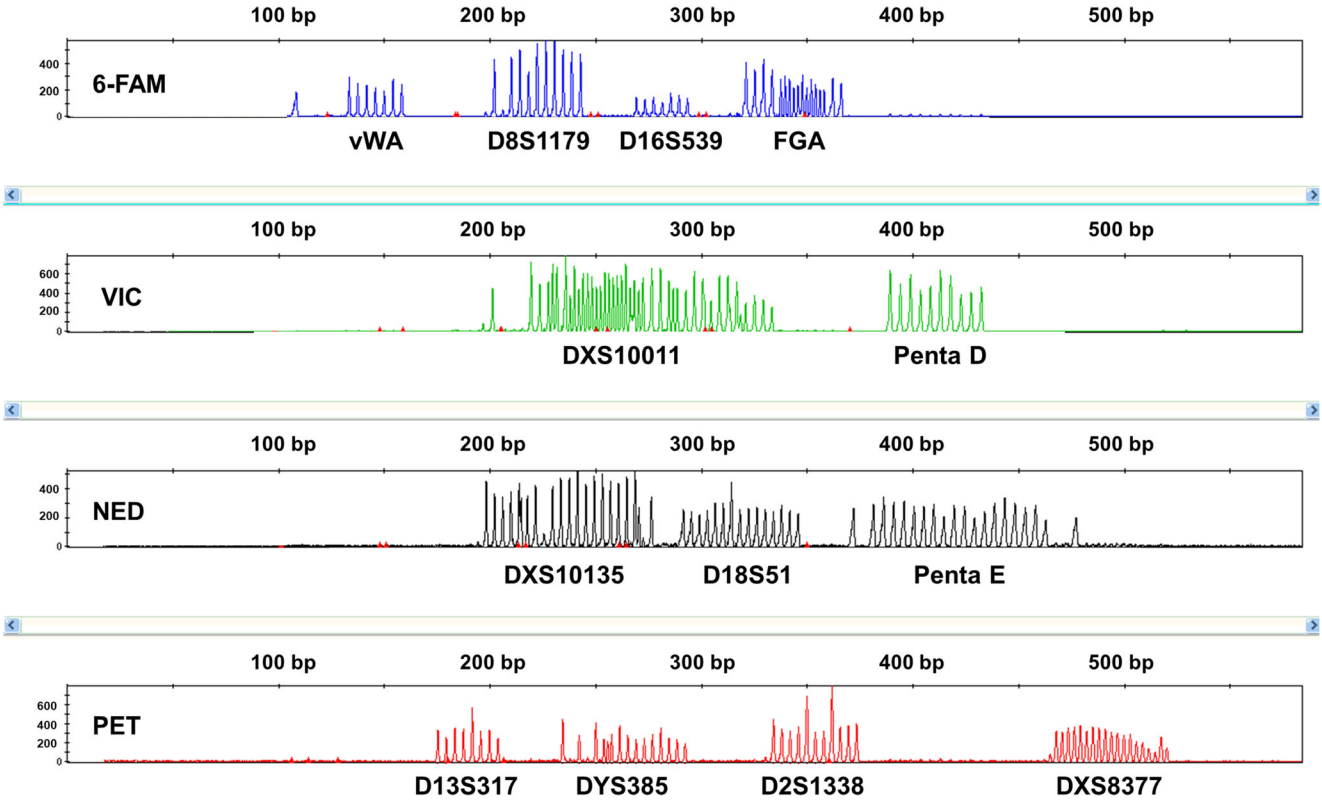

Fig. 2 Electropherogram of the allelic ladder showed distinct peaks from 202 alleles' fragments. The ladders from the left to the right represent alleles from (blue) vWA, D8S1179, and D16S539, FGA; (green) DXS10011 and Penta D; (yellow) DXS10135, D18S51, and Penta E; and lastly (red) D13S317, DYS385, D2S1338, and DXS8377

When we performed population study on 253 Malaysian individuals using the 13-loci STR panel, we have found 202 unique alleles from all 13 loci. To the best of our knowledge, this is the first population study which covers DXS10011, DXS8377, and DXS10135 for all three broad ethnic groups (Malay, Chinese, and Indian) in Malaysia. The inclusion of a Y-STR locus, the DYS385, allows simultaneous STR genotyping in addition to determination of gender without the use of Amelogenin which only serve as gender discrimination marker.

Accurate allele calling was achieved by accompanying a sequence-validated allelic ladder with each run according to the ISFH guidelines for STR analysis [27]. The sequences of each allele in the allelic ladder were determined by NGS sequencing rather than the conventional Sanger sequencing. This is because, although Sanger sequencing has the advantage of producing longer reads, it is known to face the difficulties in sequencing repetitive region which often results in abrupt, premature signal loss and noisy data. In addition, unlike Sanger sequencing which requires pure single-allele PCR product as starting material for each run, NGS sequencing allows parallel sequencing of multiple (all of the 202) alleles in a single reaction. This avoids the laborious separation and purification of heterozygous alleles as well as the time-consuming sequential sequencing of alleles in hundreds of separate runs. Deep coverage of more than 100× also ensures accurate determination of STR sequences after the sequencing [28, 29]. The NGS sequencing of 202 unique alleles revealed five previously unreported rare variant alleles which include two microvariants (41.2 and 45.1) and three large-size variants (51, 52, 53) from DXS10011. The three large-size variants were larger than the largest known allele in the STR database (STRbase and ChrX-STR). Collectively, the 13-loci STR panel showed higher combined probability of exclusion (CPE) and cumulative power of discrimination (CPD) for the three main ethnic groups (Malay, Chinese, and Indian) in Malaysia as compared to previous population study using the AmpFLSTR Identifiler kit (Table 3). The high CPD and CPE despite a fewer total number of loci in this panel as compared to the AmpFLSTR Identifiler kit suggests that the 13-loci panel may potentially serve as an efficient panel for forensic and kinship analysis in addition to chimerism analysis.

When we compared the informativeness of loci in our STR panel and the AmpFLSTR Identifiler kit, each locus from the 13-loci STR panel was informative in at least 50% of samples while some loci in the AmpFLSTR Identifiler kit were only informative for as low as 30% of samples tested. This suggests that the 13-loci STR panel exhibits more balanced loci informativity compared to the AmpFLSTR Identifiler kit. In addition, the top four most informative loci were also found to

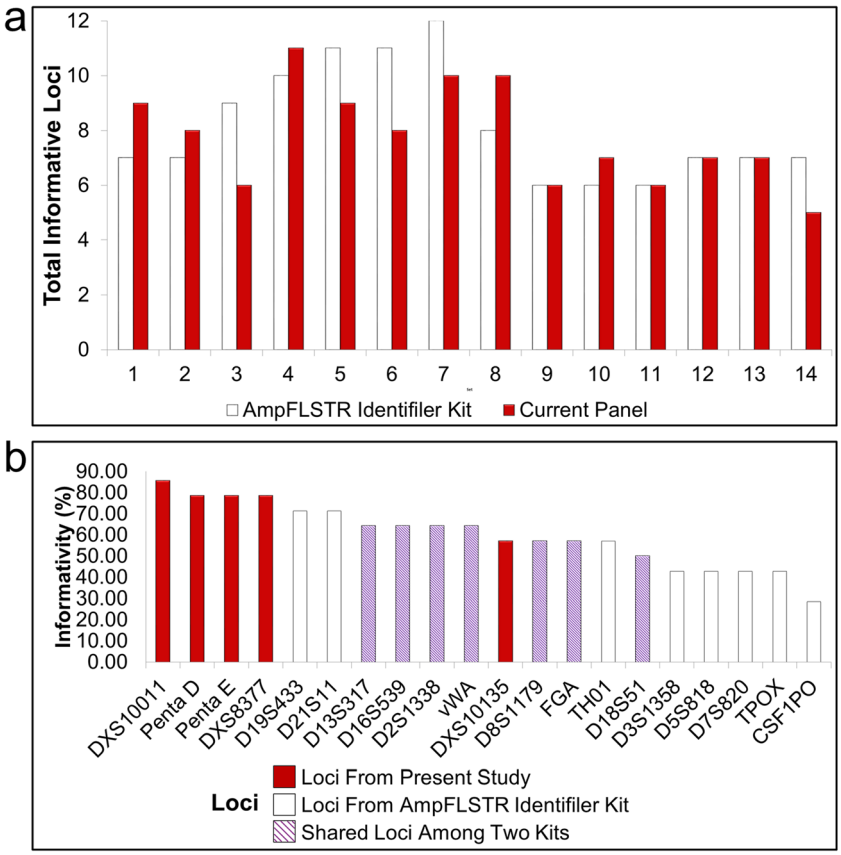

Fig. 3 **a** Comparison of total informative loci from the new 13-loci STR panel and the AmpFLSTR Identifiler kit for 14 sets of donor–recipient HSCT samples. **b** Informativity ranking of all loci from two panels. Solid-colored columns represent loci from the 13-loci STR panel, whereas line-filled columns represent loci from the AmpFLSTR Identifiler kit. Dotted-filled columns represent shared loci between two panels

be originated from our 13-loci panel, of which two are X-STRs. This suggests that although rarely used for chimerism analysis, X-STRs are useful for chimerism analysis. As for Y-STR, the DYS385 also showed high PD in Malaysia population, in agreement with previous study (Supplementary Table 1). However, as the donor and recipient pairs from UKMMC archive samples were all siblings which had inherited the same copy of Y chromosome from their father, DYS385 became non-informative in all cases and can only serve as a gender determination marker. Nevertheless, the DYS385 may potentially become a highly informative marker to discriminate unrelated donor–recipient pairs.

We have also compared the CPD and CPE of our STR panel with the AmpFLSTR Identifiler kit (Table 3). The newly developed panel generally showed a higher CPD and CPE than the AmpFLSTR Identifiler kit except in the Indian ethnic group. The lower CPE among Indian population shown in our study may be attributed by a lower sample number from the Indian ethnic group, compared to the study by Seah et al. [25]. The results indicated that our panel has a higher discriminatory power in differentiating donors and recipients and thus is more suitable in chimerism analysis in the Malaysian population.

Many commercial STR kits commonly used for chimerism analysis were originally designed for forensic identification and have not been optimized for chimerism monitoring. Chimerism monitoring requires sensitive and accurate determination of low chimeric percentage to allow early detection of relapse in order for pre-emptive therapeutic interventions to be taken [5, 30]. The sensitivity and accuracy of chimerism analysis were previously known to be affected by intra- and inter-loci balance as well as signals interference from stutter peaks [31]. The optimization of primers concentration for all loci in this study has enabled the panel to achieve good intra-loci and inter-loci balance (ratio > 0.6) which met the requirements for correct and reliable genotyping [31].

In terms of stutter percentage, all tetranucleotide STR loci in this study have achieved satisfactory stutter percentage which was lower than 16%, similar to previous study [8]. On the other hand, all trinucleotide STR loci also showed stutter percentage less than 16% except for DXS8377 (20.63%), similar to previous study [32]. The inclusion of peak area contributed by stutters may cause over-estimation of alleles' abundance. The adoption of the calculation method which deducts spilled peak area contributed by stutters has allowed accurate determination of abundance regardless of the percentage of stutters present [19]. When we analyzed a series of mocked chimerism samples, good accuracy (R^2 = 0.9968) and precision (SD < 2.49%) were achieved using the 13-loci STR panel. Minority cell population as low as 1% was detectable, but inaccuracy increased dramatically as chimerism percentage drops below 5%. Therefore, a sensitivity threshold was set at 5% for our panel, corresponding to a

CV of approximately 35% for all markers, similar to other studies [18, 33–35].

The study could help to solve the problem of STR informativity in Asian populations and their less compliance with worldwide available commercial STR kits in chimerism analysis following HSCT. However, further comparative studies are needed for the assessment of the usefulness of new kit in worldwide populations.

Conclusion

In conclusion, a multiplex STR panel containing 13 loci which are evenly high in informativeness was successfully developed for efficient chimerism analysis among HSCT patients.

Acknowledgements The authors thank Rozi Hanisa from the National Blood Centre, Malaysia, for collecting the samples in this study.

Author contributions All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Funding This study was supported by ERGS Grant (ERGS/1/2012/SKK09/UKM/01/3), Ministry of Education, Malaysia.

Compliance with ethical standards

This study was approved by the ethics committee of the UKM Medical Centre, Malaysia.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Park B, Yoo KH, Kim C (2015) Hematopoietic stem cell expansion and generation: the ways to make a breakthrough. Blood Res 50(4):194–203. https://doi.org/10.5045/br.2015.50.4.194
2. Gratwohl A, Baldomero H, Ispizua AU (2003) The EBMT activity survey on hematopoietic stem cell transplantation: a novel instrument for quality control. Turk J Med Sci 33:281–288
3. Faraci M, Bagnasco F, Leoni M et al (2018) Evaluation of chimerism dynamics after allogeneic hematopoietic stem cell transplantation in children with Non-malignant diseases. Biol Blood Marrow Transplant 24(5):1088–1093. https://doi.org/10.1016/j.bbmt.2017.12.801
4. Bader P, Niethammer D, Willasch A, Kreyenberg H, Klingebiel T (2005) How and when should we monitor chimerism after allogeneic stem cell transplantation? Bone Marrow Transplant 35(2):107–119. https://doi.org/10.1038/sj.bmt.1704715
5. Clark JR, Scott SD, Jack AL, Lee H, Mason J, Carter GI, Pearce L, Jackson T, Clouston H, Sproul A, Keen L, Molloy K, Folarin N', Whitby L, Snowden JA, Reilly JT, Barnett D (2015) Monitoring of chimerism following allogeneic haematopoietic stem cell transplantation (HSCT): technical recommendations for the use of short tandem repeat (STR) based techniques, on behalf of the United Kingdom National External Quality Assessment Service for Leucocyte Immunophenotyping Chimerism Working Group. Br J Haematol 168(1):26–37. https://doi.org/10.1111/bjh.13073
6. Lion T, Watzinger F, Preuner S et al (2012) The EuroChimerism concept for a standardized approach to chimerism analysis after allogeneic stem cell transplantation. Leukemia 26(8):1821–1828. https://doi.org/10.1038/leu.2012.66
7. Butler JM (2012) Advanced topics in forensic DNA typing: methodology. Elsevier Academic Press, San Diego
8. Butler JM (2005) Forensic DNA typing: biology, technology, and genetics of STR markers. Elsevier Academic Press, San Diego
9. Butler JM (2005) Constructing STR multiplex assays. Methods Mol Biol 297:53–66
10. Butler JM (2006) Genetics and genomics of core short tandem repeat loci used in human identity testing. J Forensic Sci 51(2):253–265. https://doi.org/10.1111/j.1556-4029.2006.00046.x
11. Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet 66(3):979–988. https://doi.org/10.1086/302825
12. Romualdi C, Balding D, Nasidze IS et al (2002) Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. Genome Res 12(4):602–612. https://doi.org/10.1101/gr.214902
13. Ariffin H, Daud SS, Mohamed Z et al (2007) Evaluation of two short tandem repeat multiplex systems for post-haematopoietic stem cell transplantation chimerism analysis. Singap Med J 48(4):333–337
14. Han E, Kim M, Kim Y, Han K, Lim J, Kang D, Lee GD, Kim JR, Lee JW, Chung NG, Cho B, Eom KS, Kim YJ, Kim HJ, Lee S, Cho SG, Min CK, Kim DW, Lee JW, Min WS (2017) Practical informativeness of short tandem repeat loci for chimerism analysis in hematopoietic stem cell transplantation. Clin Chim Acta 468:51–59. https://doi.org/10.1016/j.cca.2017.02.004
15. Saberzadeh J, Miri MR, Tabei MB, Dianatpour M, Fardaei M (2016) Genetic variations of 21 STR markers on chromosomes 13, 18, 21, X, and Y in the south Iranian population. Genet Mol Res 15(4):1–9. https://doi.org/10.4238/gmr15049065
16. Krenke BE, Tereba A, Anderson SJ et al (2002) Validation of a 16-locus fluorescent multiplex system. J Forensic Sci 47(4):773–785
17. Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF (2002) A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. Forensic Sci Int 129(1):10–24
18. Thiede C, Bornhauser M, Oelschlagel U et al (2001) Sequential monitoring of chimerism and detection of minimal residual disease after allogeneic blood stem cell transplantation (BSCT) using multiplex PCR amplification of short tandem repeat-markers. Leukemia 15(2):293–302
19. Chen DP, Tseng CP, Tsai SH et al (2008) Systematic analysis of stutters to enhance the accuracy of chimerism testing. Ann Clin Lab Sci 38(3):264–272
20. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10(3):564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x
21. Goodwin W, Linacre A, Hadi S (2007) An introduction to forensic genetics. John Wiley & Sons, New Jersey
22. Desmarais D, Zhong Y, Chakraborty R (1998) Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). J Forensic Sci 43(5):1046–1049
23. Fondevila M, Santos C, Phillips C et al (2011) An assessment of linkage between forensic markers: core STRs, mini-STRs and

Indels. Proceedings of 22nd International Symposium of Human Identification, Washington
24. Szibor R, Hering S, EdelmannJ (2006) A new web site compiling forensic chromosome X research is now online. Int J Legal Med 120(4):252–254. doi: https://doi.org/10.1007/s00414-005-0029-y
25. Seah LH, Jeevan NH, Othman MI, Jaya P, Ooi YS, Wong PC, Kee SS (2003) STR data for the AmpFlSTR Identifiler loci in three ethnic groups (Malay, Chinese, Indian) of the Malaysian population. Forensic Sci Int 138(1–3):134–137
26. van Rood JJ, Oudshoorn M (2008) Eleven million donors in bone marrow donors worldwide! Time for reassessment? Bone Marrow Transplant 41(1):1–9. https://doi.org/10.1038/sj.bmt.1705866
27. Bar W, Brinkmann B, Budowle B et al (1997) DNA recommendations. Further report of the DNA commission of the ISFH regarding the use of short tandem repeat systems. International Society for Forensic Haemogenetics. Int J Legal Med 110(4):175–176
28. Gettings KB, Kiesler KM, Fait SA et al (2016) Sequence variation of 22 autosomal STR loci detected by next generation sequencing. Forensic Sci Int Genet 21:15–21. https://doi.org/10.1016/j.fsigen.2015.11.005
29. Treangen TJ, Salzberg SL (2011) Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 13(1):36–46. https://doi.org/10.1038/nrg3117
30. Odriozola A, Riancho JA, Colorado M, Zarrabeitia MT (2013) Evaluation of the sensitivity of two recently developed STR multiplexes for the analysis of chimerism after haematopoietic stem cell transplantation. Int J Immunogenet 40(2):88–92. https://doi.org/10.1111/j.1744-313X.2012.01123.x
31. ENFSI (2010) Recommended minimum criteria for the validation of various aspects of the DNA profiling process. http://enfsi.eu/wp-content/uploads/2016/09/minimum_validation_guidelines_in_dna_profiling_-_v2010_0.pdf. Accessed 4 May 2015
32. Zarrabeitia MT, Amigo T, Sanudo C et al (2002) Sequence structure and population data of two X-linked markers: DXS7423 and DXS8377. Int J Legal Med 116(6):368–371
33. Nollet F, Billiet J, Selleslag D, Criel A (2001) Standardisation of multiplex fluorescent short tandem repeat analysis for chimerism testing. Bone Marrow Transplant 28(5):511–518. https://doi.org/10.1038/sj.bmt.1703162
34. Thiede C, Florek M, Bornhauser M et al (1999) Rapid quantification of mixed chimerism using multiplex amplification of short tandem repeat markers and fluorescence detection. Bone Marrow Transplant 23(10):1055–1060. https://doi.org/10.1038/sj.bmt.1701779
35. Nuckols JD, Rasheed BK, McGlennen RC et al (2000) Evaluation of an automated technique for assessment of marrow engraftment after allogeneic bone marrow transplantation using a commercially available kit. Am J Clin Pathol 113(1):135–140. https://doi.org/10.1309/QP7P-J49V-8Q15-36MT