IDENTIFICATION OF SARS-COV-2 LINEAGES USING A COST-EFFECTIVE MASSARRAY® SYSTEM

Bhakbhoom Panthan[1,2], Theerarat Kochakarn[1,3,4], Poramate Jiaranai[1,2], Chakkaphan Runcharoen[2], Insee Sensorn[1,2], Angkana Charoenyingwattana[2], Ekawat Pasomsub[1,4], Namfon Kotanan[1,3], Treewat Watthanachockchai[4], Kingkan Rakmanee[4], Chutatip Tongsook[4], Thanat Chookajorn[1,5] and Wasun Chantratita[1,2]

[1]COVID-19 Network Investigations Alliance (CONI); [2]Center for Medical Genomics, Faculty of Medicine Ramathibodi Hospital; [3]Genomics and Evolutionary Medicine Unit (GEM), Center of Excellence in Malaria Research, Faculty of Tropical Medicine; [4]Division of Virology, Department of Pathology, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; [5]Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Abstract. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of COVID-19. Identification of virus lineages is key to tracking the chains of transmission and for monitoring variants of interest as well as variants of concern. This study aimed to develop and validate a cost-effective method to determine SARS-CoV-2 lineages of virus samples from the first wave of the pandemic in Thailand (January - May 2020). A lineage-specific single nucleotide polymorphism (SNP) genotyping panel of SARS-CoV-2 was designed based on genomic surveillance data generated in the first wave. Viral RNA specimens obtained from the Department of Pathology, Faculty of Medicine Ramathibodi Hospital were analyzed by a MassARRAY® platform in comparison to a next generation sequencing of the virus genome. The MassARRAY® system was able to identify the SARS-CoV-2 lineages with high accuracy supporting its application as a rapid and cost-effective method for identification of circulating COVID-19 variants in the country, and has the ability to accommodate new SARS-CoV-2 variants by simple modifications of the assay protocol.

Keywords: SARS-CoV-2, coronavirus, COVID-19, MassARRAY, Thailand

Correspondence: Thanat Chookajorn, COVID-19 Network Investigations Alliance (CONI), Mahidol University, Bangkok 10400, Thailand
E-mail: thanat.cho@mahidol.edu
and
Wasun Chantratita, COVID-19 Network Investigations Alliance (CONI), Mahidol University, Bangkok 10400, Thailand
E-mail: wasun.cha@mahidol.ac.th

INTRODUCTION

Since the first report of coronavirus disease 2019 (COVID-19) in Wuhan, China in December 2019, there have been more than 130 million confirmed cases and almost 3 million deaths worldwide (as of 4 April 2021) (WHO, 2021). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the etiological agent of COVID-19 causing acute hypoxemic respiratory distress and multiple organ failure (Li *et al*, 2020; Wilcox, 2020). SARS-CoV-2 is an enveloped, single-stranded positive-strand RNA virus with a genome size of approximately 29.9 kb (Lokman *et al*, 2020). Its closest relative (96.2% nucleotide sequence homology) is a bat (Order Chiroptera) SARS-related coronavirus (SARSr-CoV; RaTG13) collected in Yunnan, PR China (Tang *et al*, 2020). Tracing virus lineages is crucial for determining how the virus spreads in communities at a population scale (Gough, 2020).

There are currently more than one million viral genomic sequences of SARS-CoV-2 submitted to the Global Initiative on Sharing Avain Influenza Data (GISAID: https://www.gisaid.org/) by clinicians and researchers from all over the world. Several systems were developed for classifying virus lineages in circulation. One of the popular classification systems is PANGO, which designates viral lineages with letters and numerical values (*eg*, lineage A.1 or lineage B.2) (Rambaut *et al*, 2020). This system is dynamic and adaptable to emerging populations. For example, a virus lineage unique to Thailand descending from an A lineage (originally designated A/Thai-1) emerged in March 2020 and later was classified as A.6 (Joonlasak *et al*, 2020).

During the course of a pandemic, whole-genome sequencing (WGS) provides the most comprehensive data of a virus genetic repertoire. However, it is relatively expensive to carry out especially in developing countries (Kiselev *et al*, 2020). In addition, WGS requires expertise both at the bench and the bioinformatic workflow to produce high-quality data and to perform comprehensive analyses (Buermans and den Dunnen, 2014). New approaches for virus genotyping are needed, which are simple and cost-effective. The Agena Bioscience MassARRAY® System is a highly sensitive tool for high-throughput single nucleotide polymorphism (SNP) genotyping and mutation profiling (Jurinke *et al*, 2002; Ellis and Ong, 2017). The system is based on the matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF), and genetic variants are distinguished by analysis of their masses (Jurinke *et al*, 2002). The MassARRAY® system is a sensitive, cost-effective and robust method with a multiplex capability (simultaneous detection of up to 40 different SNPs in a single reaction

well) and requires minimal expertise in bioinformatics to assign virus lineages.

In this study, we developed a MassARRAY®-based method to test and validate its application in SARS-CoV-2 lineage identification. A panel designed for virus lineages in circulation in Thailand was tested and validated with patients' samples in comparison to results from a WGS method. The method should assist in effectively identifying SARS-CoV-2 lineages in clinical samples with acceptable sensitivity and specificity.

MATERIALS AND METHODS

SARS-CoV-2 RNA collection

Viral RNA specimens (n = 122) were obtained from the Department of Pathology, Faculty of Medicine Ramathibodi Hospital, Mahidol University Bangkok, Thailand. The study protocol was approved by the Institutional Review Board of Ramathibodi Hospital, Mahidol university (approval no. MURA2020/676).

Lineage-defining mutations selection

We selected each nucleotide position based on genomic data produced by COVID-19 Network Investigations (CONI) alliance (https://coni.team/) and chose the position that is unique for each lineage to design the SARS-CoV-2 lineage panel (based on PANGO lineage version 7 May 2020 used during the genomic surveillance in the first wave (January - May 2020) (Joonlasak *et al*, 2020), with SARS-CoV-2 genome (MN908947.3) as standard nucleotide sequence and Assay Design Suite (ADS) v2.0 (Agena Bioscience, San Diego, CA) for developing the panel. Lineages (n = 21) were defined by 34 variants (2-3 variants per lineage) (Table 1) and their locations in the SARS-CoV-2 genome is depicted in Fig 1. A panel in one plexes (one sample per one reaction) was generated following the ADS criteria of minimizing primer to primer interactions, hairpin loops and mismatch primings. Fifty amplification primers and 34 extension primers (Supplementary Table 1 available at https://bit.ly/3ikVjzD) were synthesized by Macrogen Inc (Seoul, South Korea).

Complementary (c)DNA preparation

cDNA was generated from viral RNA using a SuperScript™ IV First-Strand Synthesis System (Invitrogen, Carlsbad, CA). In brief, RNA sample was denatured at 65°C for 5 minutes, immediately placed on ice for 1 minute and 11 µl aliquot added to a reaction mixture containing 1 µl of 50 µM random hexamers, 1 µl of 10 mM dNTPs mix, 4 µl of 5X RT buffer (Invitrogen, Carlsbad, CA), 1 µL of 0.1 M dithiothreitol, 40 U Ribonuclease Inhibitor (Invitrogen, Carlsbad, CA), and 200 U SuperScript® IV RT (Invitrogen, Carlsbad, CA), followed by incubation at 42°C for 50 minutes and 70°C for 10 minutes. Samples were

Table 1

SARS-CoV-2 lineage based on single nucleotide variant

Single nucleotide variant*	Lineage
C8782CT	A
C28144C	A
A17858G	A.1
C18060T	A.1
G25979T	A.2
C28863T	A.2
C24034T	A.3
G28077C	A.3
G28878A	A.4
G29742A	A.4
C26088T	A.5
C2942T	A/Thai-1
G20134T	A/Thai-1
G24047A	A/Thai-1
C3037T	B.1
C14408T	B.1
A23403G	B.1
A20268G	B.1.5
A24862G	B.1.8
G22427A	B.1.47
C3373A	B.1.56
C11195T	B.1.56
C14805T	B.2, A.2
G26144T	B.2, B.7
C2558T	B.2.1
C9223T	B.2.2
T17247C	B.2.2, B.2.3
G1440A	B.3

Table 1 (cont)

Single nucleotide variant*	Lineage
G2891A	B.3
G1397A	B.4
T28688C	B.4
C13730T	B.6
C28311T	B.6
C9962T	B.7

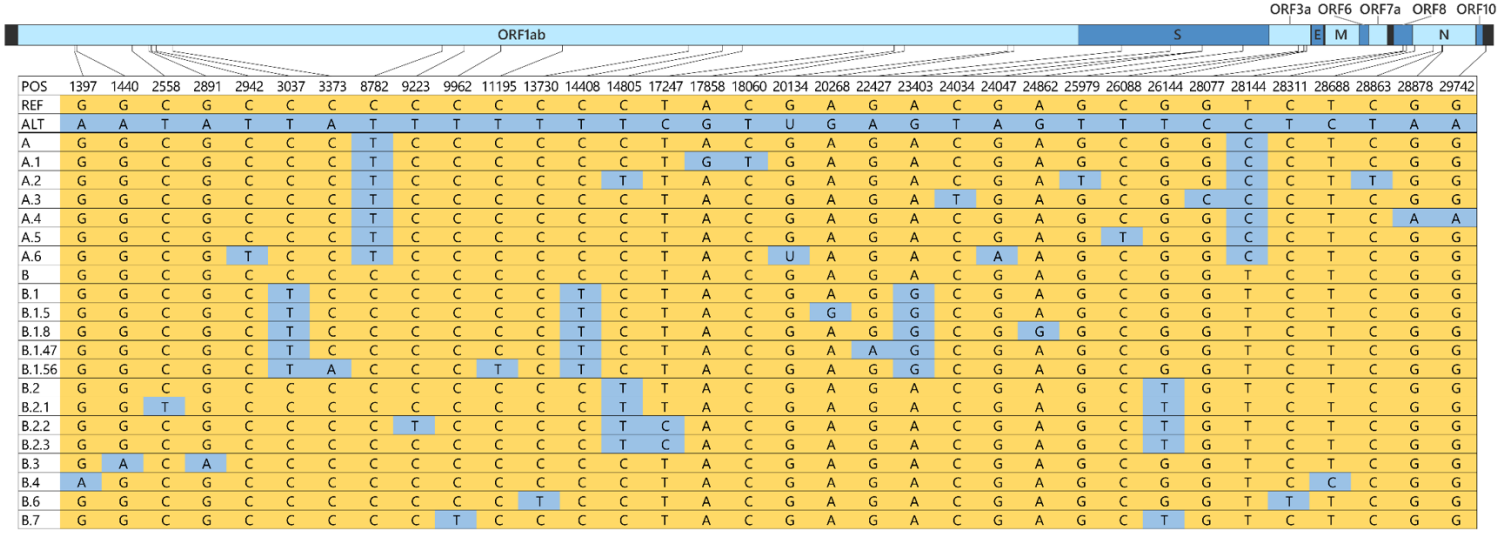
*Reference is SARS-CoV-2 GenBank accession no. MN908947.3.

SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

stored at -20°C until used.

PCR amplification protocol

PCR multiplex reactions were performed in three steps according to the manufacturer's protocol (Agena Bioscience, San Diego, CA). In step 1, the reaction solution contained 0.5 µl of 10 X PCR buffer (Agena Bioscience, San Diego, CA), 0.4 µl of 25 mM $MgCl_2$, 0.1 µl of 25 mM dNTPs, 1 U hot start Taq polymerase (Agena Bioscience, San Diego, CA), 0.5 µM each amplification primer, 2 µl of cDNA template or healthy human cDNA as a negative control or nuclease-free water as blank control, and nuclease-free water to a final volume of 5.0 µl. Thermocycling conditions, conducted in C1000 Touch™ Thermal Cycler (Bio-Rad, Munich, Germany) were as follows: 95°C for 2 minutes; 45 cycles of 95°C for 30 seconds, 56°C for 30 seconds and 72°C for 60 seconds; and a final extension step of 72°C for 5 minutes. In step 2, a 7.0 µl aliquot of a solution comprising of 0.5 U shrimp alkaline phosphatase (SAP) (Agena Bioscience, San Diego, CA) (to inhibit functioning amplification primers and dNTPs), 0.2 µl 10X SAP buffer (Agena Bioscience, San Diego, CA), and 6.5 µl nuclease-free water was added to the step 1 solution, which was then heated at 37°C for 40 minutes, followed by at 85°C for 5 minutes. In step 3, a single base extension reaction was carried out. A reaction mixture comprising of 0.2 µl of 10X iPlex buffer (Agena Bioscience, San Diego, CA), 0.2 µl of iPlex termination mix (Agena Bioscience, San Diego, CA, 1.3 U iPlex enzyme (Agena Bioscience, San Diego, CA , 0.9 µl of 5-12 µM (according to primer mass) extension primers, and 0.6 µl of nuclease-free water was added to step 2 solution. Extension reaction was performed in a thermocycler described above as follows: an initial incubation at 95°C for 30 seconds, followed by 40 cycles



POS	1397	1440	2558	2891	2942	3037	3373	8782	9223	9962	11195	13730	14408	14805	17247	17858	18060	20134	20268	22427	23403	24034	24047	24862	25979	26088	26144	28077	28144	28311	28688	28863	28878	29742
REF	G	G	C	G	C	C	C	C	C	C	C	C	C	C	T	A	C	G	A	G	A	C	G	A	G	C	G	G	T	C	T	C	G	G
ALT	A	A	T	A	T	T	A	T	T	T	T	T	T	T	C	G	T	U	G	A	G	T	A	G	T	T	T	C	C	T	C	T	A	A
A	G	G	C	G	C	C	C	T	C	C	C	C	C	C	T	A	C	G	A	G	A	C	G	A	G	C	G	G	C	C	T	C	G	G
A.1	G	G	C	G	C	C	C	T	C	C	C	C	C	C	T	G	T	G	A	G	A	C	G	A	G	C	G	G	C	C	T	C	G	G
A.2	G	G	C	G	C	C	C	T	C	C	C	C	C	T	T	A	C	G	A	G	A	C	G	A	T	C	G	G	C	C	T	T	G	G
A.3	G	G	C	G	C	C	C	T	C	C	C	C	C	C	T	A	C	G	A	G	A	T	G	A	G	C	G	C	C	C	T	C	G	G
A.4	G	G	C	G	C	C	C	T	C	C	C	C	C	C	T	A	C	G	A	G	A	C	G	A	G	C	G	G	C	C	T	C	A	A
A.5	G	G	C	G	C	C	C	T	C	C	C	C	C	C	T	A	C	G	A	G	A	C	G	A	G	T	G	G	C	C	T	C	G	G
A.6	G	G	C	G	T	C	C	T	C	C	C	C	C	C	T	A	C	U	A	G	A	C	A	A	G	C	G	G	C	C	T	C	G	G
B	G	G	C	G	C	C	C	C	C	C	C	C	C	C	T	A	C	G	A	G	A	C	G	A	G	C	G	G	T	C	T	C	G	G
B.1	G	G	C	G	C	T	C	C	C	C	C	C	T	C	T	A	C	G	A	G	G	C	G	A	G	C	G	G	T	C	T	C	G	G
B.1.5	G	G	C	G	C	T	C	C	C	C	C	C	T	C	T	A	C	G	G	G	G	C	G	A	G	C	G	G	T	C	T	C	G	G
B.1.8	G	G	C	G	C	T	C	C	C	C	C	C	T	C	T	A	C	G	A	G	G	C	G	G	G	C	G	G	T	C	T	C	G	G
B.1.47	G	G	C	G	C	T	C	C	C	C	C	C	T	C	T	A	C	G	A	A	G	C	G	A	G	C	G	G	T	C	T	C	G	G
B.1.56	G	G	C	G	C	T	A	C	C	C	T	C	T	C	T	A	C	G	A	G	G	C	G	A	G	C	G	G	T	C	T	C	G	G
B.2	G	G	C	G	C	C	C	C	C	C	C	C	C	T	T	A	C	G	A	G	A	C	G	A	G	C	T	G	T	C	T	C	G	G
B.2.1	G	G	T	G	C	C	C	C	C	C	C	C	C	T	T	A	C	G	A	G	A	C	G	A	G	C	T	G	T	C	T	C	G	G
B.2.2	G	G	C	G	C	C	C	C	T	C	C	C	C	T	C	A	C	G	A	G	A	C	G	A	G	C	T	G	T	C	T	C	G	G
B.2.3	G	G	C	G	C	C	C	C	C	C	C	C	C	T	C	A	C	G	A	G	A	C	G	A	G	C	T	G	T	C	T	C	G	G
B.3	G	A	C	A	C	C	C	C	C	C	C	C	C	C	T	A	C	G	A	G	A	C	G	A	G	C	G	G	T	C	T	C	G	G
B.4	A	G	C	G	C	C	C	C	C	C	C	C	C	C	T	A	C	G	A	G	A	C	G	A	G	C	G	G	T	C	C	C	G	G
B.6	G	G	C	G	C	C	C	C	C	C	C	T	C	C	T	A	C	G	A	G	A	C	G	A	G	C	G	G	T	T	T	C	G	G
B.7	G	G	C	G	C	C	C	C	C	T	C	C	C	C	T	A	C	G	A	G	A	C	G	A	G	C	T	G	T	C	T	C	G	G

Fig 1 - Single nucleotide variants determined by MassARRAY® in SARS-CoV-2 genomes in the Thai population (January - May 2020)

Reference SARS-CoV-2 genome sequence (GenBank accession no. MN908947.3) is shown in yellow and single nucleotide variant in blue. The variant sequence used to assign PANGO lineages are shown in each respective row.

ALT: single nucleotide variant detected; POS: position; REF: reference sequence; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

of one step at 95°C for 5 seconds with five subcycles of 52°C for 5 seconds and 80°C for 5 seconds, and final extension step at 72°C for 3 minutes.

MassARRAY® assay

A 50 µl mixture comprising of 9 µl of step 3 solution and 41 µl of nuclease-free water were desalted using ion exchange resin (Agena Bioscience, San Diego, CA) and then spotted onto a matrix pre-coated 96-SpectroCHIP® Array (Agena Bioscience, San Diego, CA) and subjected to mass spectrometry assay using a MassARRAY® MALDI-TOF MS system (Agena Bioscience, San Diego, CA) according to manufacturer's instructions. Genotyping from M/z^+ signals were analyzed using a MassARRAY® Typer v4.0 (Agena Bioscience, San Diego, CA).

Sensitivity and specificity of MassARRAY® SARS-CoV-2 lineages

Genotyping data from the MassARRAY® panel were compared with those from the next-generation sequencing (NGS) method using an Illumina's MiSeq sequencer (Illumina Inc, San Diego, CA). Sequences were then mapped to the SARS-CoV-2 reference genome (MN908947.3) and the lineage for each sample was identified using a Pangolin workflow (Center for Genomic Pathogen Surveillance, Cambridgeshire, England). PANGO lineage assignment was based on genomic sequences obtained from Illumina sequencing method.

MassARRAY® variant positions in each SARS-CoV-2 sample from Illumina sequencing data were compared with those from at each position with available data from every sample for the sensitivity and specificity calculations (Tian *et al*, 2016). Sensitivity and specificity of MassARRAY® SARS-CoV-2 genotyping are calculated as follows (Tian *et al*, 2016).

$$\text{Sensitivity} = \frac{\text{Number of variant positions identified in agreement by both MassARRAY® and NGS}}{\text{Number of variant positions identified by NGS}}$$

$$\text{Specificity} = \frac{\text{Number of variant positions in agreement between MassARRAY® and NGS}}{\text{Number of variant positions by NGS relative to the reference sequence}}$$

where NGS refers to next-generation sequencing.

RESULTS

We determined SARS-CoV-2 lineages in the Thai population during the first wave of COVID 19 epidemic (January - May 2020) using the MassARRAY® method and based on the PANGO system. From a total of 122 samples tested, seven virus lineages were identified (Table 2). The MassARRAY® data revealed 39% of the samples could be assigned to lineage A.6, 39% to B.1, 4% to B.6, 3% to B.4, 2% to B.1.5, 2% to A, and 1% to B.3. Using to Illumina sequencing data as the gold standard, sensitivity and specificity was 99 and 100% respectively (Table 3).

Among 34 single nucleotide variants (SNVs) detected by the MassARRAY® method, two positions had the lowest identity rate, namely, T18060 and A28878 (Fig 2A). This could either be the result of poor cDNA synthesis or primer bias. The rate of SNV detection by the MassARRAY® method ranged from 50-100% when NGS genomic coverage was ~100% and decreased concomitantly with drop in NGS coverage (Fig 2B). There was an inverse relationship between rate of SNV detection by the MassARRAY® method of each sample and RT-qPCR cycle threshold (Ct) values of SARS-CoV-2 ORF1ab gene employed for detecting COVID-19 status of donor samples (Fig 2C) (Syrmis *et al*, 2011). Ct value and genotyping coverage for each sample are shown in Supplementary Table 2 available at https://bit.ly/3ikVjzD.

DISCUSSION

We employed the MassARRAY® system (Jurinke *et al*, 2002) to detect SNVs in SARS-CoV-2 genomes and assign lineages. Lineages in 6% of the samples with Ct values between 30 to 40 could not be assigned due to low rate SNV detection. Other limitations of the MassARRAY® method are that the method can only detect known SARS-COV-2 variants (Yan *et al*, 2020) and success of assay is dependent on cDNA quality (Trembizki *et al*, 2014).

As of now, several sub-lineages of SARS-COV-2 B.1 have become the major virus populations in a global scale, including B.1.1.7 (Kent variant) and B.1.351 (South Africa variant) (Gough, 2020). These variants did not exist at the time of the first wave (January - May 2020), but the MassARRAY® system allows detection of any new SARS-COV-2 variants by inclusion of appropriate new primers.

MALDI-TOF MS has been used for SNP genotyping in which MS directly assesses the molecular mass of the amplicons (Ganova-Raeva and Khudyakov, 2013; Hong *et al*, 2008). The MassARRAY® system has an advantage over other MALDI-TOF MS assays by delivering a simple workflow that allows the user to customize the assay design process and to modify procedures to match circulating virus lineages. In addition, the MassARRAY® system has a broad

Table 2

SARS-CoV-2 lineages in Thai population (January - May 2020)

Name	GISAID ID	SARS-CoV-2 lineage classification	
		PANGOlearn[a] (version 30 Oct 2020)	MassARRAY®[b]
BKKCOVID17	EPI_ISL_423039	B.1.1	B.1
BKKCOVID18	EPI_ISL_450723	A.6	A.6
BKKCOVID19	EPI_ISL_423040	A.6	A.6
BKKCOVID20	EPI_ISL_423041	B.1	B.1
BKKCOVID21	EPI_ISL_423042	B.1	B.1
BKKCOVID22	EPI_ISL_423043	B.1.1	B.1
BKKCOVID25	EPI_ISL_429164	A.6	A.6
BKKCOVID26	EPI_ISL_429165	A.6	A.6
BKKCOVID28	EPI_ISL_429166	A.6	A.6
BKKCOVID30	EPI_ISL_429168	A	A.6
BKKCOVID33	EPI_ISL_429169	B.1.1	B.1
BKKCOVID34	EPI_ISL_429170	A.6	A.6
BKKCOVID35	EPI_ISL_429171	A.6	A.6
BKKCOVID36	EPI_ISL_429172	B.1.1	B.1
BKKCOVID37	EPI_ISL_429173	A.6	A.6
BKKCOVID38	EPI_ISL_429174	B.1	B.1
BKKCOVID39	EPI_ISL_429175	B.4	B.4
BKKCOVID40	EPI_ISL_429176	A.6	A.6
BKKCOVID41	EPI_ISL_429177	A.6	A.6
BKKCOVID42	EPI_ISL_429178	B.1	B.1
BKKCOVID43	EPI_ISL_429179	B.1	B.1.5
BKKCOVID44	N/A	B.1	B.1
BKKCOVID45	N/A	B.1	B.1
BKKCOVID46	N/A	B.1	B.1
BKKCOVID47	N/A	B.1	B.1
BKKCOVID48	N/A	B.1	B.1

Table 2 (cont)

Name	GISAID ID	SARS-CoV-2 lineage classification	
		PANGOlearn[a] (version 30 Oct 2020)	MassARRAY®[b]
BKKCOVID54	EPI_ISL_446996	B.1	B.1.5
BKKCOVID55	EPI_ISL_446997	B.1	B.1
BKKCOVID56	EPI_ISL_446998	B.1	B.1
BKKCOVID57	EPI_ISL_446999	B.1	B.1
BKKCOVID58	EPI_ISL_447000	B.1	B.1
BKKCOVID59	EPI_ISL_447001	B.1	B.1
BKKCOVID60	EPI_ISL_447002	A.6	A.6
BKKCOVID61	EPI_ISL_447003	B.2.1	B.2.1/B.2,A.2/B.2,B.7
BKKCOVID62	EPI_ISL_447004	B.1	B.1
BKKCOVID63	EPI_ISL_447005	A.6	A.6
BKKCOVID64	EPI_ISL_447006	B.1.1	B.1
BKKCOVID65	EPI_ISL_447007	A.6	A.6
BKKCOVID66	EPI_ISL_447008	B.1.1	B.1
BKKCOVID67	EPI_ISL_447009	B.1	B.1
BKKCOVID68	EPI_ISL_447010	B.1.1	B.1
BKKCOVID69	EPI_ISL_447011	B.3	B.3
BKKCOVID70	EPI_ISL_447012	B.1	B.1
BKKCOVID72	EPI_ISL_447013	B.1	B.1
BKKCOVID73	EPI_ISL_447014	B.1	B.1
BKKCOVID74	EPI_ISL_447015	B.1.1	B.1
BKKCOVID75	EPI_ISL_447016	B.1	B.1
BKKCOVID76	EPI_ISL_447017	B.1	B.1
BKKCOVID77	EPI_ISL_447018	B.1	B.1
BKKCOVID78	EPI_ISL_447019	B.4	B.4
BKKCOVID79	EPI_ISL_447020	A.6	A.6
BKKCOVID80	EPI_ISL_447021	A.6	A.6
BKKCOVID81	EPI_ISL_447022	B.1	B.1

Table 2 (cont)

Name	GISAID ID	SARS-CoV-2 lineage classification	
		PANGOlearn[a] (version 30 Oct 2020)	MassARRAY®[b]
BKKCOVID82	EPI_ISL_447023	B.1	B.1
BKKCOVID83	EPI_ISL_447024	A.6	A.6
BKKCOVID84	EPI_ISL_447025	B.4	B.4
BKKCOVID85	EPI_ISL_447026	A.6	A.6
BKKCOVID86	N/A	B.1	B.1
BKKCOVID87	EPI_ISL_447027	B.1	B.1
BKKCOVID88	EPI_ISL_447028	B.1.1.7	B.1/B.1.56/B.2,A.2
BKKCOVID89	EPI_ISL_447029	A.6	A.6
BKKCOVID91	N/A	A.6	A
BKKCOVID92	N/A	A.6	A
BKKCOVID94	EPI_ISL_458023	B.1	B.1
BKKCOVID95	EPI_ISL_458024	B.1.5	B.1
BKKCOVID96	EPI_ISL_458025	B.1	B.1
BKKCOVID97	EPI_ISL_458026	B.1	B.1
BKKCOVID98	EPI_ISL_458027	B.1	B.1
BKKCOVID99	EPI_ISL_458028	B.1.1	B.1
BKKCOVID105	EPI_ISL_455908	A.6	A.6
BKKCOVID106	EPI_ISL_455909	A.6	A.6
BKKCOVID107	EPI_ISL_455910	B.6	B.6
BKKCOVID108	EPI_ISL_455911	A.6	A.6
BKKCOVID109	EPI_ISL_455912	A.6	A.6
BKKCOVID110	EPI_ISL_455913	B.4	B.4
BKKCOVID111	EPI_ISL_455914	A.6	A.6
BKKCOVID112	N/A	B.1.1	B.1
BKKCOVID113	EPI_ISL_455915	B.1.1.10	B.1
BKKCOVID114	EPI_ISL_455916	A.6	A.6
BKKCOVID115	EPI_ISL_455917	B.6	B.6

Table 2 (cont)

Name	GISAID ID	SARS-CoV-2 lineage classification	
		PANGOlearn[a] (version 30 Oct 2020)	MassARRAY®[b]
BKKCOVID116	EPI_ISL_455918	B.1.1	B.1
BKKCOVID117	EPI_ISL_455919	B.6	B.6
BKKCOVID126	EPI_ISL_455920	B.1	B.1.5
BKKCOVID127	EPI_ISL_455921	B.6	B.6
BKKCOVID128	EPI_ISL_455922	B.2	B.2.2,B.2.3/B.2,A.2/ B.2,B.7
BKKCOVID129	EPI_ISL_455923	B.1	B.1
BKKCOVID130	EPI_ISL_455924	A.6	A.6
BKKCOVID131	EPI_ISL_455925	A.6	A.6
BKKCOVID132	EPI_ISL_455926	A.6	A.6
BKKCOVID133	EPI_ISL_455927	B.6	B.6
BKKCOVID134	EPI_ISL_455928	A.6	A.6
BKKCOVID135	EPI_ISL_455929	A.6	A.6
BKKCOVID136	EPI_ISL_455930	A.6	A.6
BKKCOVID137	EPI_ISL_455931	A.6	A.6
BKKCOVID138	EPI_ISL_455932	A.6	A.6
BKKCOVID139	EPI_ISL_455933	A.6	A.6
BKKCOVID140	EPI_ISL_455934	B.1.1	B.1
BKKCOVID141	EPI_ISL_455935	A.6	A.6
BKKCOVID142	EPI_ISL_455936	A.6	A.6
BKKCOVID143	EPI_ISL_455937	A.6	A.6
BKKCOVID144	EPI_ISL_455938	A.6	A.6
BKKCOVID145	EPI_ISL_455939	A.6	A.6
BKKCOVID146	EPI_ISL_455940	A.6	A.6
BKKCOVID147	EPI_ISL_455941	A.6	A.6
BKKCOVID148	EPI_ISL_455942	A.6	A.6

Table 2 (cont)

Name	GISAID ID	SARS-CoV-2 lineage classification	
		PANGOlearn[a] (version 30 Oct 2020)	MassARRAY®[b]
BKKCOVID149	EPI_ISL_455943	B.1	B.1
BKKCOVID150	EPI_ISL_455944	A.6	A.6
BKKCOVID151	N/A	B.1	B.1
BKKCOVID152	EPI_ISL_455945	A.6	A.6
BKKCOVID153	EPI_ISL_455946	A.6	A.6
BKKCOVID154	EPI_ISL_455947	A.6	A.6
BKKCOVID155	EPI_ISL_455948	A.6	A.6

[a]Rambaut *et al* (2020); [b]Jurinke *et al* (2002)

GISAID: Global Initiative on Sharing Avain Influenza Data; ID: Identity Document; N/A: not assigned; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Table 3

Comparison of SARS-CoV-2 single nucleotide variants obtained from MassARRAY® method and next generation sequencing

MassARRAY® method[a] (number of positions)	Next generation sequencing[b] (number of positions)		Total
	Single nucleotide variant	Wild type nucleotide	
Single nucleotide variant	377	0	377
Wild type nucleotide	2	2,822	2,824
Total	379	2,822	3,201

[a]Agena Bioscience, San Diego, CA; [b]Illumina Inc, San Diego, CA

SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

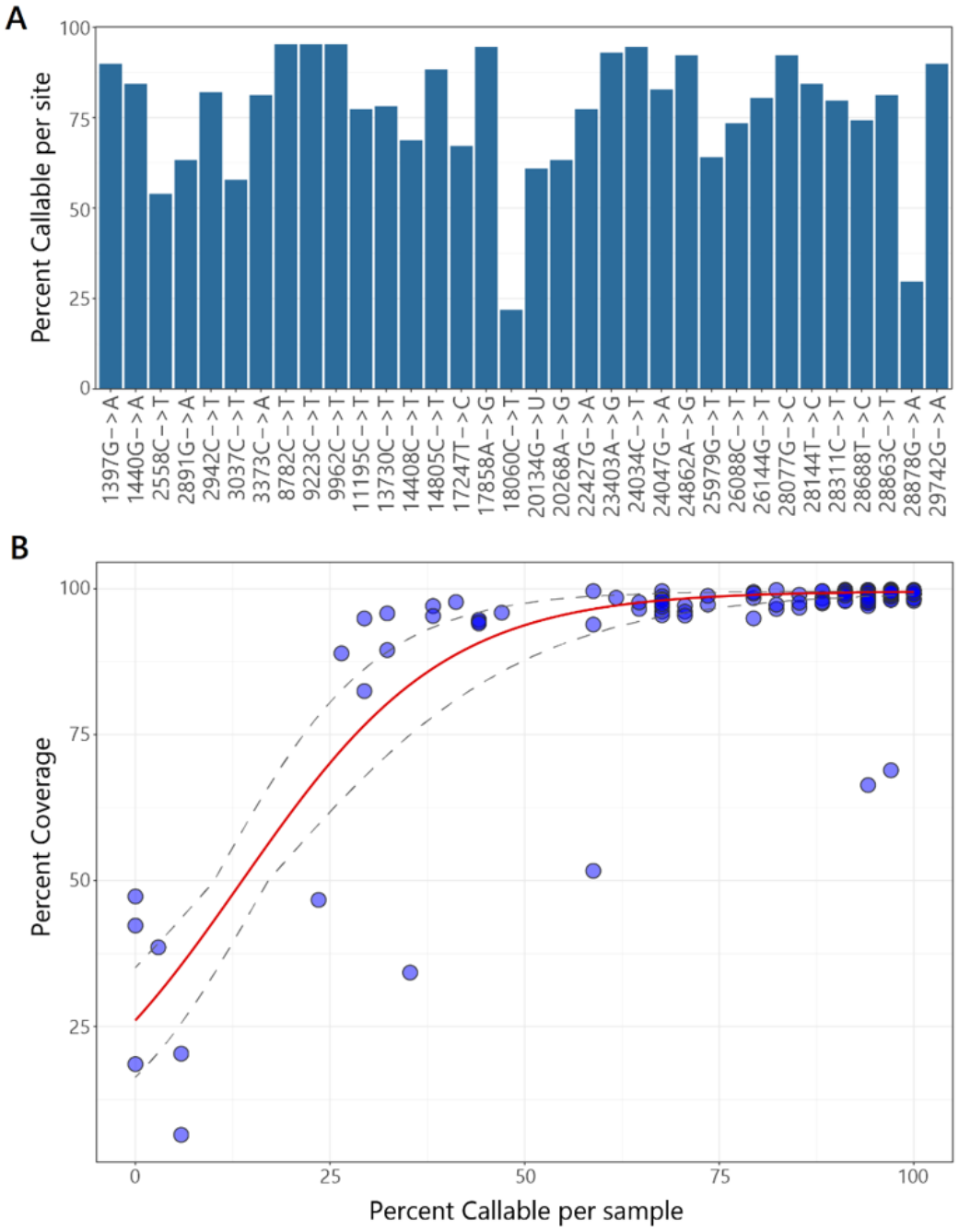


Fig 2 - Analysis of SARS-CoV-2 single nucleotide variants (SNVs) identification by MassARRAY® method

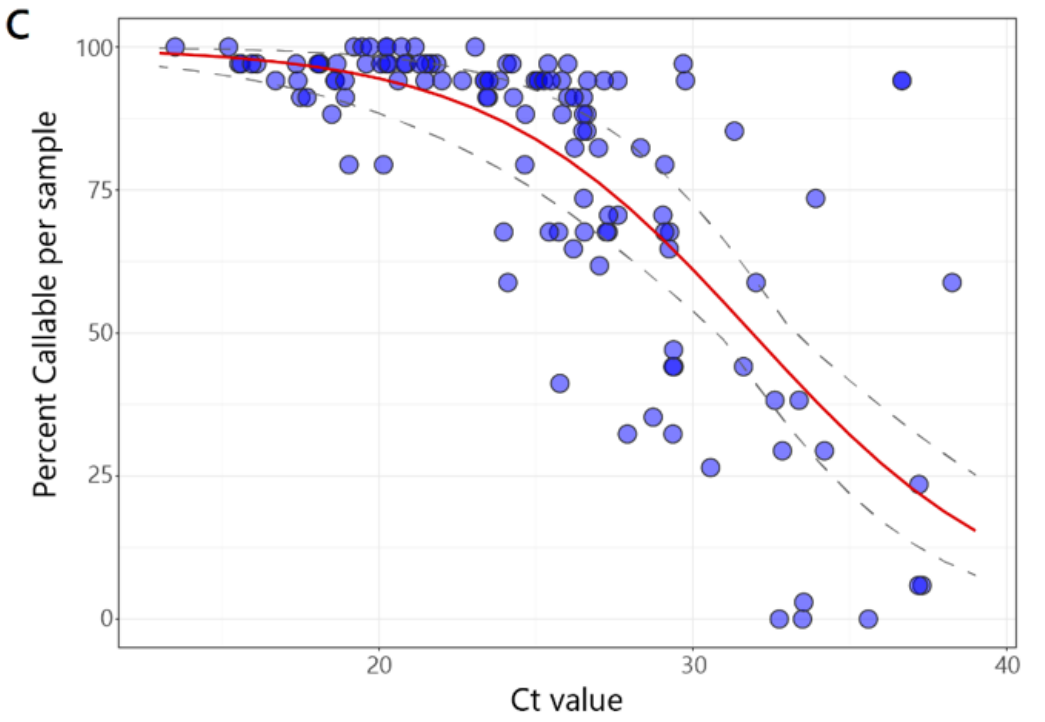


Fig 2 - (cont) Analysis of SARS-CoV-2 single nucleotide variants (SNVs) identification by MassARRAY® method

A: Percent samples with SNVs determined by MassARRAY® method relative to SARS-CoV-2 GenBank accession no. MN908947.3 reference sequence; B: Comparison between percent callable (identifiable) SNVs by MassARRAY® method and next generation sequencing (Illumina Inc, San Diego, CA); C: Comparison between percent callable SNVs by MassARRAY® and RT-qPCR cycle threshold (Ct) values [(*ORF1ab* gene using SARS-CoV-2 Nucleic Acid Detection Kit (Zybio Inc, Chongqing, PR China)] (Syrmis *et al*, 2011).

Blue dot represents data point from an individual sample. Nonlinear least squares fit using Gompertz function is shown in red line and 95% confidence interval in grey dotted lines.

RT-qPCR: quantitative real-time reverse transcriptase polymerase chain reaction; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; SNV: single-nucleotide variant

level of multiplexing resulting in significant reduction in genotyping cost (Ragoussis *et al*, 2006).

In conclusion, we report here the utility of MassARRAY® system in identifying circulating SARS-CoV-2 lineages with high accuracy and ease of operation. Rapid identification of SARS-CoV-2 lineages will assist in formulating appropriate the public health responses to current and emerging COVID-19 pandemics.

ACKNOWLEDGEMENTS

The study was supported by Ramathibodi Foundation. The authors thank all participants who kindly assisted in the study.

CONFLICTS OF INTEREST DISCLOSURE

The authors declare no conflicts of interest.

REFERENCES

Buermans HPJ, den Dunnen JT. Next generation sequencing technology: advances and applications. *Biochim Biophys Acta* 2014; 1842: 1932-41.

Ellis JA, Ong B. The MassARRAY® system for targeted SNP genotyping. *Methods Mol Biol* 2017; 1492: 77-94.

Ganova-Raeva LM, Khudyakov YE. Application of mass spectrometry to molecular diagnostics of viral infections. *Expert Rev Mol Diagn* 2013; 13: 377-88.

Gough NR. Defining a new strain of a virus, 2020 [cited 2021 Apr 28]. Available from: URL: https://www.bioserendipity.com/defining-a-new-strain-of-a-virus/

Hong SP, Shin SK, Lee EH, *et al*. High-resolution human papillomavirus genotyping by MALDI-TOF mass spectrometry. *Nat Protoc* 2008; 3: 1476-84.

Joonlasak K, Batty EM, Kochakarn T, *et al*. Genomic surveillance of SARS-CoV-2 in Thailand reveals mixed imported populations, a local lineage expansion and a virus with truncated ORF7a. *Virus Res* 2021; 292: 198233.

Jurinke C, van den Boom D, Cantor CR, Koster H. The use of MassARRAY technology for high throughput genotyping. *Adv Biochem Eng Biotechnol* 2002; 77: 57-74.

Kiselev D, Matsvay A, Abramov I, Dedkov V, Shipulin G, Khafizov K. Current trends in diagnostics of viral infections of unknown etiology. *Viruses* 2020; 12: 211.

Li Q, Wu J, Nie J, *et al*. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* 2020; 182: 1284-94.e9.

Lokman SM, Rasheduzzaman M, Salauddin A, *et al*. Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: a computational biology approach. *Infect Genet Evol* 2020; 84: 104389.

Ragoussis J, Elvidge GP, Kaur K, Colella S. Matrix-assisted laser desorption/

ionisation, time-of-flight mass spectrometry in genomics research. *PLoS Genet* 2006; 2: e100.

Rambaut A, Holmes EC, O'Toole A, *et al*. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 2020; 5: 1403-7.

Syrmis M, Moser R, Whiley DM, *et al*. Comparison of a multiplexed MassARRAY system with real-time allele-specific PCR technology for genotyping of methicillin-resistant Staphylococcus aureus. *Clin Microbiol Infect* 2011; 17: 1804-10.

Tang X, Wu C, Li X, *et al*. On the origin and continuing evolution of SARS-CoV-2. *Natl Sci Rev* 2020; 7: 1012-23.

Tian HX, Zhangn XC, Wang Z, *et al*. Establishment and application of a multiplex genetic mutation-detection method of lung cancer based on MassARRAY platform. *Cancer Biol Med* 2016; 13: 68-76.

Trembizki E, Smith H, Lahra M, *et al*. High-throughput informative single nucleotide polymorphism-based typing of Neisseria gonorrhoeae using the Sequenom MassARRAY iPLEX platform. *J Antimicrob Chemother* 2014; 69: 1526-32.

Wilcox SR. Management of respiratory failure due to COVID-19. *BMJ* 2020; 369: m1786.

World Health Organization (WHO). COVID-19 weekly epidemiological update, 6 April 2021, 2021 [cited 2021 Apr 20]. Available from: URL: https://apps.who.int/iris/handle/10665/340695

Yan Y, Jin X, Wang X, *et al*. Screening of PAH common mutations in Chinese phenylketonuria patients using iPLEX MALDI-TOF MS. *ACS Omega* 2020; 5: 1805-12.